Thesis/
Reports
Kluzek,
E.

FINAL REPORT FOR RESEARCH AGREEMENT
#INT-94883-RJVA
"Synthesis of a Weather Sequence Climate
Generator Based on a Mesoscale Model"
UTAH STATE UNIVERSITY
FS Contact: Dr. William J. Elliot
CoOp Contact: Dr. David Bowels

Development of Mountain Climate Generator and Snowpack Model for Erosion Predictions in the Western United States using WEPP

Final Report – Climate Modeling and Data Analysis

Research Joint Venture Agreement INT-94883-RJVA

Submitted to:

U.S. Department of Agriculture, Forest Service
Rocky Mountain Research Station
Forestry Sciences Laboratory
1221 South Main Street
Moscow, Idaho 83843

Submitted by:
Erik Kluzek, Gail E. Bingham and David S. Bowles

September 20, 1999

UTAH WATER RESEARCH LABORATORY
UTAH STATE UNIVERSITY
Logan, Utah 84322-8200

EXECUTIVE SUMMARY

This report summarizes the results of an aggressive effort to develop a procedure for generating representative historical and synthetic climate sequences at ungauged locations throughout the mountainous Western United states. A significant problem in managing mountainous areas is the lack of measured climate data available to the resource manager, and the inability to extrapolate existing data due to terrain variation. The objective of this effort was to determine if meso scale modeling could improve data availability by utilizing the extensive data sets collected daily for weather prediction to supplement or replace the more traditional data sets. The Utah Water research Laboratory (UWRL) conducted this project under a research joint venture agreement with the U.S. Forest Service (USFS) as part of the USFS Watershed and Erosion Prediction Project (WEPP). Climate sequence simulations and software have been provided to the USFS under previous reports. This report summarizes the analysis.

Historical modeling of precipitation for long-term studies in mesoscale atmospheric models has proven to be a challenging task. We present results using a mesoscale atmospheric physics model at meso-gamma resolutions. To develop high-resolution experiments we first developed a 50-kilometer simulation over the western United States for a 14-year period. The output of the 50-kilometer simulation was used for 10-kilometer resolution experiments. For the purpose of this study we turned to a higher resolution meso-gamma-scale limited area model at 10-kilometer resolution nested within a meso-beta-scale limited area model at 50-kilometer resolution. First we developed methodology to compare observations with the regional climate model simulations at either resolution. Next we examined a 14-year simulation at the 50-kilometer scale. And finally, we used the output from the 50-kilometer simulation as input for a 10-kilometer nested simulation. We did this for a specific month within the longer simulation.

We examined some of the difficulties with modeling precipitation in this type of setup. One concern we examined closely was the conservation of mass and heat when going from the meso-beta-scale model to meso-gamma-scale. The problems we considered are over prediction of precipitation in the 14 year, 50 Km runs and the distribution of precipitation over mountainous regions in the high resolution runs. Our high resolution simulations overall had much too little precipitation, but unrealistically high amounts were clustered on the windward sides of mountain slopes. Our findings suggested that much of the difficulty with modeling at this scale has to do with the following issues: the explicit cloud microphysics, the quality of the input 50-kilometer simulation, its update frequency, the domain size, and finally the need for cloud and rain water to be input from the 50-kilometer simulation.

We conclude that while driving today's mesoscale models with global scale meteorological inputs provide monthly averaged regional scale data of significant quality, it does not provide local scale data that is an improvement over a data using surface data extrapolation. Thus, it is probably not in the interest of a Forest Service modeling project studying the details of either erosion or plant and animal growth to invest in a mesoscale modeling effort at this point. The "too wet" and "too often" convective precipitation phenomenon that exists in RegCM2 after all our effort strongly suggests that locally accurate precipitation sequences cannot be generated by " more carefully parameterized" hydrostatic simulations. The new generation regional models becoming available allow for observed rainfall and temperature to be injected as an internal grid boundary condition input to more closely constrain storm phasing and surface layer variables. However the value of

this approach is limited to historical exploration and short term forecasts. While computer power and data storage capability has increased by an order of magnitude in the last 5 years, the requirements of long term simulations with non-hydrostatic codes still exceed the computer capabilities of Forest Service sized research centers. The effort-accuracy winner still lies with an advanced surface observation extrapolation scheme, such as the local polynomial scheme utilized in Chapter II.

ACKNOWLEDGMENTS

This work was partially funded through the Forest Service Intermountain Research Station, Moscow, ID under joint contract INT-92669-RJVA.

TABLE OF CONTENTS

EXECUTIVE SUMMARY ii
TABLE OF CONTENTS iv
LIST OF TABLES vi
LIST OF FIGURES vii
LIST OF ACRONYMS xii
GLOSSARY xiv
LIST OF SYMBOLS xix

CHAPTER I: INTRODUCTION 1

CHAPTER II: DOWNSCALING GCM SCALE OUTPUT TO MOUNTAINOUS REGIONS USING A REGIONAL MODEL: PART I: COMPARISON METHODOLOGY 6

II.1 Introduction 6
II.2 Discussion 7
II.2.1 Observed Datasets 8
II.2.2 Adjustments to Model Data 8
II.3 Comparison Methodology 10
II.3.1 Comparison Regions 11
II.3.2 Observed Data Gridding Methods 11
II.3.3 Observed Data Error Checking 11
II.3.4 Neighborhood Definition 12
II.3.5 Observed Data Gridding Techniques 12
II.3.6 Local Polynomial Regression 13
II.3.7 Average Estimate 14
II.3.8 Regional Average Method 15
II.3.9 Comparison Statistics 16
II.3.10 Climatology 16
II.3.11 Threat Score, Bias Score, and Monte-Carlo Threat Score 17
II.3.12 Precipitation Frequency 18
II.4 Results 19
II.4.1 Observed Data Error Checking Method 19
II.4.2 Difference in Error Estimates 20
II.4.3 Difference in Climatology 20
II.5 Discussion and Conclusions 21

CHAPTER III: DOWN SCALING GCM SCALE OUTPUT TO MOUNTAINOUS REGIONS USING A REGIONAL MODEL: PART II: LONG-TERM STATISTICAL COMPARISON 33

III.1 Introduction 33
III.2 Study Areas 35
III.2.1 Regional Atmospheric Model: RegCM2 35
III.3 Results 37
III.3.1 Model and Observed Climatology 37

III.3.2 Temperature and Precipitation Forecast Skill 39
III.3.3 Precipitation Frequencies 41
III.3.4 Precipitation Intensity Distribution 42
III.3.5 Wet and Dry Spell Analysis 42
III.4 Discussion and Conclusions 43

CHAPTER IV: THE IMPACT OF LATERAL BOUNDARY CONDITIONS ON REGIONAL CLIMATE MODELING: A HIGH RESOLUTION TEST CASE IN COMPLEX TERRAIN 85

IV.1 Introduction 85
IV.2 Discussion 86
IV.2.1 Issues with High Resolution and Modeling of Cumulus Systems 88
IV.2.2 Quality of the Input LBC Data to the 10 km RegCM2 Simulations 88
IV.2.3 ECMWF Dataset Compared to Observed 89
IV.2.4 Simulation Input Dataset Compared to Observed 89
IV.2.5 Simulation Input Dataset Compared to the ECMWF Dataset 89
IV.2.6 Theoretical Mechanism for Artificially Enhanced LBC Storms 90
IV.3 Results 91
IV.4 Discussion and Conclusions 94

CHAPTER V: SUMMARY AND CONCLUSIONS 126

LIST OF TABLES

Table 1. List of Regions and Abbreviations Used for Each Region 25

Table 2. Difference Between the Average Elevation for the Model Grid-Points Within Each Region and the Average of the Observed Station Elevations for Eleven (250 km)^2 Mountainous Regions. The Second Column is the Average of All the Model Grid-Points Within the Region. The Third Column is the Average of the Elevations for the Scattered Observation Stations Within the Region. And Finally, the Last Column is the Difference From Model to Observed 25

Table 3. Suite of Comparision Statistics 26

Table 4. Impact of the Data Screening on Precipitation Estimates over the Entire 14-year Period 27

LIST OF FIGURES

Figure 1. Eleven regions we use for comparison. 28

Figure 2. Location of the observed stations and grid points for six of the eleven analysis areas. Green circles are the station locations and red x's are the grid points. 29

Figure 3. Location of the observed stations and grid points for five of the eleven analysis areas. Green circles are the station locations and red x's are the grid points. 30

Figure 4. Comparison of T_{ave} error estimates. 31

Figure 5. Comparison of error estimates for precipitation. 31

Figure 6. Period averages of average temperature for the AE, LP, and AVG analysis methods. 32

Figure 7. Period averages of monthly precipitation for the AE, LP, and AVG analysis methods. 32

Figure 8. RegCM2 50 km terrain contour map of the western US. Compare these smoothed profiles with the DEM generated terrain in Figure 1 50

Figure 9. Maximum monthly temperature computed using the model and observed climatologies with standard deviations for six of the eleven analysis areas. 51

Figure 10. Maximum monthly temperature computed using the model and observed climatologies with standard deviations for five of the eleven analysis areas. 52

Figure 11. Plot of model and observed Minimum monthly temperature climatologies, with standard deviations for six of the eleven analysis areas. 53

Figure 12. Plot of model and observed Minimum monthly temperature climatologies, with standard deviations for five of the eleven analysis areas. 54

Figure 13. Mean monthly temperature computed using the model and observed climatologies with standard deviations for six of the eleven analysis areas. 55

Figure 14. Mean monthly temperature computed using the model and observed climatologies with standard deviations for five of the eleven analysis areas. 56

Figure 15. Monthly precipitation computed using the model and observed climatologies with standard deviations for six of the eleven analysis areas. 57

Figure 16. Monthly precipitation computed using the model and observed climatologies with standard deviations for five of the eleven analysis areas. 58

Figure 17. Cumulative monthly precipitation computed using the model and observed climatologies for six of the eleven analysis areas. 59

Figure 18. Cumulative monthly precipitation computed using the model and observed climatologies for five of the eleven analysis areas. 60

Figure 19. Annual average observed error for each region. 61

Figure 20. Model absolute bias between the model and observed climatologies. 61

Figure 21. Correlation between the model and observed monthly anomalies. 62

Figure 22. Model predictive skill compared with a predictive model based only on the observed climatology 62

Figure 23. . Maximum air temperature model skill scores developed using a simple model based on the bias corrected observed climatology (perfect skill = 1). These plots are smoothed using a 30-day running mean. 63

Figure 24. Maximum air temperature model skill scores developed using a simple model based on the bias corrected observed climatology (perfect skill = 1). The plots are smoothed using a 30-day running mean. 64

Figure 25. The model precipitation skill score. When the skill score is 1, modeled and observed data agree perfectly. At a skill score of zero, the model is predicting precipitation at the same level as the climatology. At a skill level of –1, the model is only half as good as the climatology predictor is. 65

Figure 26. The model precipitation skill score. When the skill score is 1, modeled and observed data agree perfectly. At a skill score of zero, the model is predicting precipitation at the same level as the climatology. At a skill level of –1, the model is only half as good as the climatology predictor is. 66

Figure 27. Precipitation bias score generated using daily data smoothed by a 30 day running mean. The bias score is the model/observed ratio of events of a given magnitude averaged over the 25 points for each day..... 67

Figure 28. Precipitation bias score generated using daily data smoothed by a 30 day running mean. The bias score is the model/observed ratio of events of a given magnitude averaged over the 25 points for each day..... 68

Figure 29. . Precipitation threat score analysis of the model output, compared with the observed data. When TS = 1, the model is perfect. 69

Figure 30. Precipitation threat score analysis of the model output, compared with the observed data. When TS = 1, the model is perfect. 70

Figure 31. Precipitation threat score analysis of the model output, compared with the Monte Carlo sampling of the observed data. When $TS_{mc} > 0$, the model has predictive capability. 71

Figure 32. Precipitation threat score analysis of the model output, compared with the Monte Carlo sampling of the observed data. When $TS_{mc} > 0$, the model has predictive capability. 72

Figure 33. Annual average precipitation bias score. 73

Figure 34. Annual average precipitation threat score..... 73

Figure 35. Annual average precipitation Monte-Carlo threat score..... 74

Figure 36. Winter time (December, January, and February) precipitation intensity distribution calculated for the 14-year period of study. Cumulated frequency is normalized to 1. 75

Figure 37. Winter time (December, January, and February) precipitation intensity distribution calcualted for the 14-year period of study. Cumulated frequency is normalized to 1. 76

Figure 38. Summer precipitation intensity for six of the eleven analysis areas in the western US. Cumulated frequency is normalized to 1. 77

Figure 39. Summer precipitation intensity for five of the eleven analysis areas in the western US. Cumulated frequency is normalized to 1 78

Figure 40. The wet spell length distribution comparison for the observed and model for the winter period (December, January, and February). Cummulated frequency is normalized to 1. 79

Figure 41. The wet spell length distribution comparison for the observed and model for the winter period (December, January, and February). Cummulated frequency is normalized to 1. 80

Figure 42. The wet spell length distribution comparison for the observed and model for summer period (June, July, and August). Cummulated frequency is normalized to 1. 81

Figure 43. The wet spell length distribution comparison for the observed and model for summer period (June, July, and August). Cummulated frequency is normalized to 1.. 82

Figure 44. The dry spell length distribution comparison for the observed and model for summer period (June, July, and August). Cummulated frequency is normalized to 1. 83

Figure 45. The dry spell length distribution comparison for the observed and model for summer period (June, July, and August). Cummulated frequency is normalized to 1. 84

Figure 46. Surface heights for the 10-km model domain..... 98

Figure 47. Observed precipitation over model domain for September 1981. Monthly station data was gridded to the model grid using the local polynomial regression technique..... 99

Figure 48. Observed precipitation error over model domain for September 1981. Observed errors are calculated using the local polynomial regression technique. 100

Figure 49. Locations of the observed stations and model grid points for the 10 km domain. Grid points are the green circles and the stations are the red x's..... 101

Figure 50. Accumulated precipitation from the RegCM2 model control simulation for September 1981..... 102

Figure 51. Average and RMS differences (SLC - ECMWF) from ECMWF dataset to SLC rawinsonde, as well as standard deviations for temperature. The ECMWF data was first interpolated to the 10km model domain and then the grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" ECMWF dataset standard deviation. "C", RMS difference between ECMWF and SLC. "D", average difference between ECMWF and SLC. 103

Figure 52. Average and RMS differences (SLC - ECMWF) from ECMWF dataset to SLC rawinsonde, as well as standard deviations for water vapor mixing ratio. The ECMWF data was first interpolated to the 10km model domain and then the grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" ECMWF dataset standard deviation. "C", RMS difference between ECMWF and SLC. "D", average difference between ECMWF and SLC. 104

Figure 53. Average and RMS differences (SLC - ECMWF) from ECMWF dataset to SLC rawinsonde, as well as standard deviations for relative humidity. The ECMWF data was first interpolated to the 10km model domain and then the grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" ECMWF dataset standard deviation. "C", RMS difference between ECMWF and SLC. "D", average difference between ECMWF and SLC. 105

Figure 54. Average and RMS differences (SLC - ECMWF) from ECMWF dataset to SLC rawinsonde, as well as standard deviations for wind speed. The ECMWF data was first interpolated to the 10km model domain and then the grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" ECMWF dataset standard deviation. "C", RMS difference between ECMWF and SLC. "D", average difference between ECMWF and SLC. 106

Figure 55. Average and RMS differences (SLC - input) from input dataset to SLC rawinsonde, as well as standard deviations for temperature. The grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and SLC. "D", average difference between input and SLC. 107

Figure 56. Average and RMS differences (SLC - input) from input dataset to SLC rawinsonde, as well as standard deviations for water vapor mixing ratio. The grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and SLC. "D", average difference between input and SLC. 108

Figure 57. Average and RMS differences (SLC - input) from input dataset to SLC rawinsonde, as well as standard deviations for relative humidity. The grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and SLC. "D", average difference between input and SLC. 109

Figure 58. Average and RMS differences (SLC - input) from input dataset to SLC rawinsonde, as well as standard deviations for wind speed. The grid point closest to SLC was used to compare to the rawinsonde. "A", Salt

Lake City rawinsonde dataset standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and SLC. "D", average difference between input and SLC. 110

Figure 59. Contour plot of the RMS difference between input dataset and observed daily T_{ave} for September, 1981. Observed temperature calculated using the local polynomial regression technique on the mean of daily maximum and minimum temperatures. Input dataset temperature is calculated by finding the mean of the 6 hour data. 111

Figure 60. Average observed temperature error for September 1981. Observed error calculated using the local polynomial regression technique 112

Figure 61. Average and RMS differences (ECMWF - input) of domain averaged temperature between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF. 113

Figure 62. Average and RMS differences (ECMWF - input) of domain averaged water vapor mixing ratio between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF. 114

Figure 63. Average and RMS differences (ECMWF - input) of domain averaged relative humidity between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF. 115

Figure 64. Average and RMS differences (ECMWF - input) of domain averaged wind speed between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF. 116

Figure 65. Average and RMS differences (ECMWF - input) of domain averaged total water mass between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF. 117

Figure 66. Control simulation total column cloud water totaled over September 1981 in mm of water per unit area. 118

Figure 67. Comparison of domain averaged precipitation with time. Observed (black), average of 50 km simulation over the region (green), and average of the 10 km simulation (red). Dotted grid marks each day, and the dashed grid marks each 6 hour interval 119

Figure 68. Average and RMS differences (new - old) between the domain average for the control input dataset and the new interpolation methods for temperature, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. "A", "new" standard deviation. "B" "old" dataset standard deviation. "C", RMS difference between "new" and "old". "D", average difference between "new" and "old". 120

Figure 69. Average and RMS differences (new - old) between the domain average for the control input dataset and the new interpolation methods for water vapor mixing ratio, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. "A", "new" standard deviation. "B" "old" dataset standard deviation. "C", RMS difference between "new" and "old". "D", average difference between "new" and "old". 121

Figure 70. Average and RMS differences (new - old) between the domain average for the control input dataset and the new interpolation methods for relative humidity, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. "A", "new" standard deviation. "B" "old" dataset standard deviation. "C", RMS difference between "new" and "old". "D", average difference between "new" and "old". 122

Figure 71. Average and RMS differences (new - old) between the domain average for the control input dataset and the new interpolation methods for wind speed, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. "A", "new" standard deviation. "B" "old" dataset standard deviation. "C", RMS difference between "new" and "old". "D", average difference between "new" and "old". 123

Figure 72. Precipitation for 10 km simulation with hourly LBC updates. 124

Figure 73. Comparison of grid-scale average precipitation from control to 1 hour update case. Dotted grid marks each day, and the dashed grid marks each 6 hour interval. 125

LIST OF ACRONYMS

AE	Average estimate. This is an observed gridding technique that compensates for an elevation bias, then does a simple average over the observed stations. This method compensates for the elevation bias by computing a linear regression with elevation for the observed data. After computing the regression the data-points are adjusted to the elevation desired (using this regression) and averaged.
AVG	This is a technique to obtain estimate of the average response in a region by averaging all the observed data stations within the region.
BATS	Biosphere Atmosphere Transfer Scheme. This is a surface physics and hydrology model developed by the National Center for Atmospheric Research for use in climate models.
BATS1E	The latest version of Biosphere Atmosphere Transfer Scheme (version 1E). This is also the version of Biosphere Atmosphere Transfer Scheme used in Regional Climate Model and Regional Climate Model version 2.
CAC	Climate Analysis Center.
CCM	Community Climate Model. This is a Global Climate Model developed at National Center for Atmospheric Research.
CCM2	Version 2 of the Community Climate Model.
CDF	Cumulative distribution function. This is the integral of a Probability Density Function that starts at zero and goes to 1.0 at infinity.
DEM	Digital Elevation Model.
ECMWF	European Centre of Medium Range Weather Forecasting. The center that supplies the synoptic scale observational analyses that are input into the higher resolution Regional Climate Model version 2.
ENSO	El-Nino Southern Oscillation, the oscillation in surface pressure between South America and Australia. This is an atmospheric phenomena associated with the El-Nino anomalous warming of southern Pacific Sea Surface Temperature off the pacific coast of South America. The atmosphere responds to El-Nino by a weakening of the low-latitude trade winds and a reversal of the pressure difference between South America and Australia. These changes in the global circulation also affect the climate in other regions across the globe.
GCM	Global Climate Model. This is an atmospheric model that covers the entire globe.
GMT	Greenwich Mean Time.
LAM	Limited Area Model. This is an atmospheric model that does not cover the entire globe. It requires input from either a Global Climate Model or observational data.
LBC	Lateral Boundary Conditions. The boundary conditions for a Limited Area Model are on the outside edges of the horizontal grid domain.
LGCV	Local General Cross Validation. This is a statistic used in the estimation of a value at a particular point based on a set of observations. This statistic considers the average influence of excluded observation for the predictive squared error of the estimate.
LP	Local Polynomial regression. This is an observation estimation technique that uses observed data at scattered locations to create estimates at given locations.
MCC	Mesoscale Convective Complexes.

MM4	Mesoscale Model version 4. This is a 3-dimensional hydrostatic meteorological Limited Area Model that solves the continuity, momentum, and energy equations. It also has parameterizations for surface physics, hydrology, radiation, the atmospheric Planetary Boundary Layer, cumulus clouds, and rainfall.
MM5	Mesoscale Model version 5. This is an atmospheric model, similar to Mesoscale Model version 4. This model has advanced micro-scale physics, a complex nesting scheme, hydrodynamic option, and a data inclusion scheme. It is mainly used for forecasting or short duration runs and lacks sophisticated Planetary Boundary Layer, radiation, or surface physics packages.
NCAR	National Center for Atmospheric Research.
NOAA COOP	National Oceanic and Atmospheric Administration Cooperative observation network. This is a network of stations that report daily minimum and maximum temperatures as well as precipitation.
NWP	Numerical Weather Prediction model. These are atmospheric models used for weather forecasting. Climate models are essentially the same as these. The only difference is that climate models require more accurate physics for surface, hydrology, and solar radiation.
PBL	Planetary Boundary Layer. The atmospheric layer near the surface that is dominated by turbulent eddy fluxes.
PDF	Probability Density Function. This is a statistical term. The integral of a Probability Density Function goes to one.
PMF	Probability Mass Function. This is the probability that a given precipitation intensity will occur.
RegCM	Regional Climate Model. This was developed at the National Center for Atmospheric Research and is used for regional climate modeling simulations.
RegCM2	Regional Climate Model version 2. A version of Mesoscale Model version 4 as modified by Filippo Giorgi and Gary Bates of the National Center for Atmospheric Research. It is used for long-term climate modeling simulations.
SNOTEL	SNow TELemetry network. A system of remote climate stations put into operation by the US Department of Agriculture.
SST	Sea Surface Temperature.
SVD	Singular Value Decompostion. A technique for solving systems of linear equations.
T42	This is the triangular spectral truncation of a spectral Global Climate Model to 42 waves. Triangular specifies a particular relationship between the spherical harmonic coefficients used in the model.

GLOSSARY

Accretion — This is a term from cloud physics. Accretion is when formed raindrops increase in size by merging with cloud-droplets. See also auto-conversion.

Adiabatic — A process by which heat is neither gained nor lost.

Advection — This is a term in the energy equation. It is the dot-product of a wind vector with the gradient of a quantity (for example temperature, mixing ratio, or wind)

Anomaly — This is the difference between the observed value of a quantity and the long-term average of the quantity. For example, the monthly temperature anomaly is the difference between the observed monthly temperature and the long-term average temperature for the same month.

Anomaly Correlation — This is defined as the correlation coefficient between the model and observed anomalies.

Auto- — This is the conversion of cloud-droplets into raindrops. See also conversion accretion.

Boundary Layer — This is a turbulent layer that forms in a viscous fluid when it interfaces with a solid object or another fluid. In our case our reference to the Boundary Layer always applies to the planetary boundary layer (PBL) the boundary layer formed near the surface of the earth. Because, the PBL consists of small scale eddies, the numerical model cannot explicitly solve the physics. Therefore, numerical models parameterize the PBL physics with empirical formulations.

Boundary Relaxation Scheme — Limited Area Models commonly work by over-specification of the Lateral Boundary Conditions. Boundary conditions are provided not only on the outermost domain grid-points, but also on several grid-points near the boundaries. The model then uses a mixture of its own internal physics and the Lateral Boundary Conditions to solve for values near the outermost edge. On the outermost edge the Lateral Boundary Condition values are used, and in the inner domain only the model physics is used.

Climate Change — This is a change in the long-term weather patterns. Usually this refers to global climate change or the global warming of the climate due to the increases in greenhouse gases.

Climatology — This is the long-term averages of a given season over many years. For example, the monthly climatology is the average over several years for each individual month in the year. Daily climatology is the average over many years for each individual day in a year. Climatology is taken over as long of a period as possible. In this case, when we refer to climatology, we refer to the period average of our simulation.

Closure — The relationships between the large-scale grid variables and the sub-grid parameterization of convective precipitation are not explicitly defined. Closure assumptions must be made to complete the equations mathematically. Experiments have been preformed to provide an empirical basis for different closure assumptions.

Cloud water — This is the mass mixing ratio of liquid cloud water to dry air. Cloud water is liquid water droplets small enough to remain buoyant with updrafts. Once the droplets increase enough in size, they no longer remain buoyant and fall to the ground as rain water. See also explicit cloud water and implicit cloud water.

Convective Precipitation — This is precipitation that is formed in cumulus type clouds by strong updrafts in unstable air (such as thunderstorms). Since the scale of convection is on the order of 1 to 10 kilometers, most climate models do not resolve convective systems and must parameterize their impact with methods that diagnose the large scale environment.

Continuity Equation — This is the expression for conservation of mass.

Convective Parameterization This is the parameterization in the climate model to handle convective precipitation, or precipitation from small-scale cumulonibus clouds that are formed by strong updrafts in unstable conditions.

Cumulonimbus Cumulonimbus clouds are very deep clouds formed by strong updrafts due to powerful convection. They commonly extend up to the tropopause and produce intense precipitation.

Cumulus Parameterization See Convective parameterization.

Diabatic This is the process by which heat is gained or lost. Heating due to short wave radiation and heat loss due to the emmission of long wave radiation are examples of diabatic processes.

Davies Nudging This is a specific boundary relaxation scheme.

Diffusion This is a non-physical term used in weather models to control nonlinear instabilities and reduce noise. Diffusion is applied as a second-order gradient on the quantity that diffusion acts on. In Regional Climate Model version two, both a second-order and a fourth-order gradient are used.

Distributed Ecosystem/Hydrologic Models
These are models of vegetation and streamflow that need climate inputs on a regular grid. Typically the grids need to be 100 m or smaller for such models.

Domain This is the area over which a Limited Area Model is applied. Typically this refers to the horizontal area on the globe that the model is applied over.

Downscaling This is a method for taking Global Climate Model data and scaling it to higher resolution. See also empirically based downscaling and physically based downscaling.

Dry-spell This is the amount of time that passes between precipitation events. Typically this is used in the context of a dry-spell frequency or the probability that a given dry-spell length will occur. See also wet-spell.

El-Nino This is an anomalous warming of surface Pacific Ocean temperatures off the pacific coast that can lead to extreme weather events.

Empirically Downscaling This is a downscaling method based on experimentally derived based relationships between Global Climate Model data and observed data.

Empirical Transfer Functions
This is an infinite series of functions that are solved for in order to explain variability in a dataset.

Explicit This is the method whereby cloud water and rain water are

Cloud water calculated by solving the continuity equation. In this case the accretion and auto-conversion terms are loss terms for cloud water and source terms for rain water.

Global Climate Change This is climate change occurring on a global scale. It is primarily due to an increase in greenhouse gases.

Greenhouse Gases These are gases that are clear in the visible range, but opaque in the infrared range. These gases allow solar radiation to pass through, but less long-wave radiation to pass through without first warming the atmosphere. The greenhouse gases important for climate change are: CO_2, N_2O, CFCs, O_3, and CH_4.

Grid These are regularly spaced sampling points in a map coordinate system.. The grid in Regional Climate Model version 2 is on the Lambert conformal conical map projection system.

Gridding Gridding is the process of taking observed data at irregularly spaced points to a regularly spaced grid. See also Grid.

Grid-point Storms These are numerical simulation problems observed in Regional Climate Model simulations by Giorgi. These storms create unrealistically high amounts of precipitation for a single grid-point in the model. One way to handle this problem is to increase diffusion rates.

Grid-scale This is precipitation that forms when an entire grid-cell reaches

Resolvable Precipitation saturation. The excess moisture then becomes precipitation.

Hydrostatic A hydrostatic model assumes the vertical motions in the atmosphere are in equilibrium between the air pressure and gravity.

Hydrodynamic See non-hydrostatic.

Implicit This is the method whereby cloud water is calculated by diagnosing

Cloud water the water vapor mixing ratio.

Kernal Functions These are functions used for estimation of probability density They center at an observed point and taper to zero away from that.

Lambert A map representation that is useful for displaying mid-

Conformal Conic lattitude areas, since shapes are accurately displayed. Two standard parallels are used where the map is most accurate.

Projection Lapse Rate The relationship with elevation for a variable is called the lapse rate. This term is usually applied to temperature - with a 6.5 degrees Celsius drop in temperature per 1000 meters rise in elevation being commonly used.

Latent Heat Flux This is the heat exchange between surface and atmosphere due to the evaporation of surface water into the atmosphere.

Maximum Air Temperature This is the maximum daily temperature taken at shelter height. For monthly data this is the average over the month of the values for each day.

Minimum Air Temperature This is the minimum daily temperature taken at shelter height. For monthly data this is the average over the month of the values for each day.

Mesoscale This is atmospheric phenomena that is on a scale of 10 - 200 kilometers.

Meso-α This is mesoscale phenomena that is on a scale of 100 - 200 kilometers.

Meso-β This is mesoscale phenomena that is on a scale of 20 - 100 kilometers.

Meso-γ This is mesoscale phenomena that is on a scale of 2 - 20 kilometers.

Navier Stokes Equations These are the equations used for solving for momentum of an viscious, incompressible fluid.

Nested This is a Limited Area Model domain that lies fully inside another Limited Area Model domain, typically the nested domain is also at a higher resolution.

Non-hydrostatic Non-hydrostatic models are models that use the Navier Stokes equations and explicitly solve the momentum equation in the vertical, rather than using the hydrostatic assumption (see hydrostatic).

Non-local Boundary Layer The boundary layer mixes air within the surface to the top of the layer. Most boundary layer parameterizations are local – only data near a given location are considered. In a non-local formulation, data near the surface and the top of the layer are also considered in the mixing. See also boundary layer and PBL.

Orographic This is that which is influenced by the surface elevation (orography). For example, precipitation is influenced by orography, with precipitation being increased with elevation in mid-latitudes.

Orography Orography is surface elevation.

Perturbed Climate This Climate is created by some change from current conditions. Typically this refers to the change caused by increased greenhouse gases.

Physically based Downscaling This is a downscaling method based on physical principles that uses formulas from first-principle physics. Regional modeling is considered a physically based downscaling method.

Precipitable Water This is the total amount of water in the atmospheric column.

Rain Water This is the mass mixing ratio of rain water droplets to a mass of dry air. See also cloud water.

Rawinsonde This is an experimental measurement method using weather balloons to get information on the atmospheric profile at a given location. Typically rawinsondes are flown twice a day at major airports.

Relative Humidity Relative humidity is the mass ratio of water vapor currently in the air, relative to the mass of water vapor in the air at saturation. For example, air at 100% relative humidity is at saturation.

Relaxation Region This is the region near the lateral boundaries where the model uses both internal physics and specified boundary values. See Boundary Relaxation Scheme.

Resampling This is a method of generating a synthetic data-sequence by resampling observed data. By statistically establishing relationships in the data, the same relationships can be established in the output synthetic sequence.

Screen height This is the height at which typical weather observations of temperature are taken. This range is 1.2 to 1.5 meters.

Sensible Heat Flux This is the heat exchange between the surface and atmosphere, neglecting the component of heat that goes into phase changes. See also latent heat flux.

Shelter Height See Screen Height.

Shelter Temperature This is the temperature taken from observations at screen height.

Sigma Layer This is a vertical coordinate commonly used in atmospheric models. Near the surface it follows the terrain, but at high levels it approximates a constant pressure surface.

Skin Temperature This is the temperature of the surface as "seen" from the perspective of radiation in the atmosphere. In Regional Climate Model version 2, it is the area–weighted average of the surface temperature and vegetation temperature.

Specific Humidity This is the ratio of the mass of water vapor in the atmosphere to the mass of moist air.

Spectra GCM Spectral Global Climate Model. This is a model that solves the basic dynamics of the Global Climate Model using spherical harmonics.

Spin-up The time it takes for the variables within a climate model to reach an equilibrium state is called spin-up. Atmospheric spin-up takes a few days. However, surface quantities, especially soil temperatures, take several years to reach an equilibrium condition.

Sub-Grid-Scale Precipitation See convective precipitation.

Surface Boundary Layer The part of the boundary layer nearest to the surface. In this layer the surface fluxes of latent and sensible heat are constant.

Surface Fluxes These are surface heat exchange fluxes. See latent and sensible heat fluxes.

Synoptic These are atmospheric phenomena that are on a scale of 200-1000 kilometers.

Water Vapor Mixing Ratio — The ratio of the mass of water vapor to the mass of dry air is the water vapor mixing ratio.

Wet-spell — This is the amount of time that passes during a precipitation event. Typically this is used in the context of a wet-spell frequency or the probability that a given wet-spell length will occur. See also dry-spell.

LIST OF SYMBOLS

BS	Bias score. Ratio of the number of grid-cells exceeding the given threshold of the model to the observed.
C_D	Drag coefficient, unitless.
C_{DN}	Drag coefficient under neutral stability, unitless.
C_p	Specific heat at constant pressure for dry air, 1004 J/(kg*K).
d'	Number of coefficients fit in the local polynomial estimation.
d_i	Deviation from the median observed value.
$d_{l,l}$	Distance from observed point x_i to estimation point x_l.
D_l	Set of observation points within the k nearest neighborhood of the estimate point, x_l.
e_i	Measurement error or noise from the underlying function describing the data fitted.
err_i	Estimated error of the estimate for a given grid-point.
f	Wet/dry spell or precipitation intensity values.
f_i	An observed wet/dry spell or precipitation intensity value.
F	Number of points correctly forecast for the given precipitation threshold (th).
$f(x_i)$	Underlying functional relationship between observations and location vectors x_i.
g	Gravitational acceleration (9.8), m/sec^2.
h	Bandwidth used for the kernel functions in estimating frequency distributions.
H	Horizontal resolution of the grid used for estimation, km.
k	Number of observed points to use in making the grid-point estimate.
K	Von Karmen coefficient (0.378), unitless.
K(u)	Weight function of observed point based on distance from estimation point.
K(t)	Kernel function used for estimating frequency distributions.
k_1	Rate coefficient for auto-conversion term (10^{-3}), m^{-1}.
lat	Northward geographical latitude, Degrees.
long	Eastward geographical longitude, Degrees.
m	Model data series.
m_{clim}	Period average for each month of model data series.
M_d	Median deviation of observed data from the median observation.
M	Number of model grid-points that exceed the given precipitation threshold (th).
M_r	Number of dimensions for the location vector to use in grid-point estimates. Three in our case: latitude, longitude, and elevation.
M_y	Median observed value.
n	Number of observations.
n_{min}	Minimum number of observations to use in grid-point estimates.
O	Number of points where the observed exceed the threshold (th).
o	Observed data series.
o_{clim}	Period average for each month of observed data series.
o_{rand}	Random observed value for the same month as the desired date, but different than the same day.
p	Atmospheric air pressure, hPa.
p_m	Pressure at lowest model layer, hPa.
p_{top}	Air pressure at top model layer, hPa.
p_s	Air pressure at model surface layer, hPa.
p_{sat}	Water vapor saturation pressure, hPa.
p*	Model pressure depth ($p_s - p_{top}$), hPa.
P_{RC}	Auto-conversion, rate of conversion from cloud water to rain water, $((kg/kg)sec^{-1})$.
q	Mass mixing ratio for water vapor, kg/kg.
q_c	Mass mixing ratio for cloud water, kg/kg.
q_{cth}	Threshold level for auto-conversion term, where cloud water begins to convert into rain water (0.41), g/kg.
q_r	Mass mixing ratio for rain water, kg/kg.
q_{sat}	Water vapor mixing ratio at saturation, kg/kg.
R	Ideal gas constant for dry air (287), J/(kg*K).
r_i	Ratio of deviation (d_i) of a given observation from the median (M_y) to the median deviation (M_d).
RI_b	Bulk Richardson's number. Measure of stability or instability in atmosphere, unitless.

t	Ratio of deviation from observed value to the kernel bandwidth.
th	Precipitation threshold used for calculating BS and TS, mm.
T	Atmospheric air temperature, K.
T_{veg}	Foliage temperature, K.
T_{ave}	Average of T_{max} and T_{min}, C.
T_m	Lowest model layer temperature, K.
T_{max}	Maximum daily temperature, C.
T_{min}	Minimum daily temperature, C.
T_{skin}	Skin temperature, the surface temperature "seen" from a perspective of atmospheric radiation, K.
T_{soil}	Surface temperature over bare soil, K.
TS	Threat score. Statistic that measures how good the forecast precipitation patterns matches the observed. One is a perfect score, unitless.
TS_{mc}	Monte-Carlo threat score, developed by computing the threat score and subtracting out the threat score of a random signal. Greater than zero, the threat score has predictive skill, less than or near zero, the threat score is no better than a random forecast pattern, unitless.
T_v	Virtual temperature. Temperature including the temperature increase that would occur if all the latent heat from water vapor were released, K.
u^*	Friction velocity, m/sec.
u_m	Wind speed at lowest model layer (plus "convective" component 0.1), m/sec.
W_l	Weight matrix to weight observed data in order to make estimation.
X	Data matrix of observations for each location dimension.
x_i	Location vector for a given observed (latitude, longitude, elevation).
x_l	Location vector for estimation point.
y_l	Estimated response of the grid-point.
y_i	Observed value associated with location grid-point x_i.
y_i^*	Observed values adjusted for the elevation difference between the observed point and the estimated point.
Y_l	Estimated response vector for each grid-point.
Z_m	Height above the surface of the lowest model layer, m.
Z_{sh}	Height above the surface that screen temperature observations are taken (1.5), m.
Z_{snow}	Snow roughness height, m.
Z_{soil}	Soil roughness height, m.
Z_{veg}	Vegetation roughness height, varies depending on the type of vegetation in the given model point, m.
Z_l	Matrix formed by augmenting the data matrix X with columns that represent the polynomial expansion of X to the given order of the polynomial.
α	Power relating an adiabatic exchange between temperature and pressure, $(-R/C_p)$, unitless.
β_l	Estimates of the coefficients for the estimation polynomial.
θ	Potential temperature, temperature given by adiabatic adjustment of air temperature at a given pressure to surface pressure, K.
θ^*	Sensible heat flux, $Jm^{-2}sec^{-1}$.
θ_m	Potential temperature of lowest model layer, K.
θ_{sh}	Potential temperature at screen height (Z_{sh}), K.
θ_v	Potential virtual temperature, K.
ρ	Air density, kg/m^3.
σ	Sigma layer. Terrain-following vertical coordinate $((p-p_{top})/p^*)$, unitless - Valid in 0-1.
σ_{dev}	Standard deviation.
σ_{snow}	Fraction of model grid-point covered by snow.
σ_{soil}	Fraction of model grid-point bare of snow or vegetation.
σ_{veg}	Fraction of model grid-point covered by vegetation.
Ψ	Dependence relationship of drag coefficient with stability, unitless.
$\Delta\theta_v$	Difference in potential virtual temperature from lowest model layer to the surface, K.

CHAPTER I: INTRODUCTION

Historical modeling of precipitation, especially when applied over large areas in complex terrain in long-term studies, using mesoscale limited area models (LAMs) has proven a challenging task. Extended range studies (greater than 10 days in length) are rare and practically nonexistent at higher resolutions [*Sass and Christensen*, 1995]. However, several multi-year meso-beta-scale LAM runs now exist at resolutions around 50 km [*Giorgi,* 1990; *Giorgi and Marinucci* 1991; *Dickenson et al.* 1993]. The analyses of these runs have shown promise. For the purpose of this work we look at regional 50 km model simulations input into nested 10 km simulations. The primary attention is on verifying the 50 km simulation with observations so that the simulation can be used as input into the higher resolution 10 km simulation. There are three issues we deal with: methods of observational validation of regional climate model simulations (at either 10 km or 50 km resolution), actual validation of our 50 km simulation, and finally, issues of relevance for the 10 km simulation. This constitutes chapters II, III, and IV, respectively.

The LAM we use for our study is Regional Climate Model version 2 (RegCM2) [*Giorgi et al.*, 1993a, 1993b]. RegCM2 is a model of the physics of the atmosphere using the basic Navier-Stokes equations to solve for the atmospheric state. RegCM2 is a hydrostatic, compressible, sigma layer, primitive equation atmospheric model. The basic model is Mesoscale Model version 4 (MM4) as described by *Anthes et al.* [1987]. It also has sophisticated physics packages for surface physics (Biosphere Atmosphere Transfer Scheme version 1e (BATS1E)) [*Dickinson et al.*, 1993], lakes [*Hostetler and Giorgi*, 1992], and a non-local atmospheric boundary layer [*Holtslag et al.*, 1990; *Holtslag and Boville*, 1992]. A delta-eddington approach to solar radiation [*Briegleb*, 1992] and an option to use the Grell convective parameterization scheme [*Grell et al.*, 1994] are also used with an *Arakawa and Shubert* [1974] closure. RegCM2 is an improved version of the RegCM model used in several earlier studies [*Giorgi and Bates*, 1989; *Giorgi*, 1990, 1991; *Medina,* 1991; *Giorgi et al.*, 1992; *Matthews et al.,* 1992].

To save CPU time RegCM2 uses a split-explicit time differencing approach [*Madala*, 1981] rather than the explicit time-integration approach of Brown-Campana described in *Anthes et al.* [1987]. This approach allows use of a time-step more than twice as large without loss of accuracy compared with the Brown-Campana technique. Also the pressure gradient formulations of *Corby et al.* [1972] are used to reduce numerical errors in the pressure gradient term along sloping sigma-surfaces. In this study, two model domains are used, a 50 km model domain and a nested 10 km domain. The first model domain is similar to the area over the western US studied by *Giorgi et al.* [1992] except that a grid domain of 72 x 60, 50 km square boxes, with 17 vertical layers is used. We apply terrain smoothing on the lateral boundaries, similar to *Hong and Juang* [1998]. The lowest model layer is approximately 20 m off the surface and the model top is 80 hPa. In the second model domain the grid size is 36 x 37, 10 km grid-squares with 20 vertical layers. Like the 50 km simulation we apply terrain smoothing on the lateral boundaries.

As pointed out by *Sass and Christensen* [1995], the quality of a LAM simulation is dependent not only on the quality of the model itself, but also on the quality of the input fields. Our 10 km simulations use the 50 km simulation as input data. Therefore, one of the first tasks is to compare the 50 km simulation to observations, and to create a comparison methodology that can be used for both model resolutions. The first objective of this study is to develop validation methods for regional model simulations. Next we will compare the 50 km model simulation to observations.

And, finally, we examine the 10 km simulation domain and perform several experiments. The 10 km simulations turned out to be quite challenging, and we were not able to run simulations as long as the 50 km model. In either case the 50 km simulation analysis proved to be quite important in working with the 10 km model.

In chapter II we develop methods of observational validation for regional climate models. We apply local polynomial (LP) regression [*Lall and Bosworth,* 1994], which is developed for point estimation to compute areal-average surface temperature and precipitation data at model grid-points. We also do rigorous statistical comparisons of this method to two other methods. The first method we compare uses a linear regression in elevation to take out any bias in elevation, and then does a simple average. The second ignores the elevation bias and simply averages the data in the region. We also develop a suite of statistics for use in comparing the regional climate model to the observed data. Furthermore, to make the surface temperature comparisons valid we develop the methodology to adjust model temperatures at the lowest model layer to observation heights, using boundary layer physics. This method takes the RegCM2 (BATS1E) parameterization for the surface drag coefficient, assumes the surface flux of sensible heat is constant, and derives the resultant expression for temperature as a function of height from the surface.

Chapter III expands on chapter II by applying the methodology created for model analysis to the 14-year RegCM2 50 km simulation. The model data are compared to observed data using the LP analysis method developed in chapter II. The comparison statistics developed in chapter II are also applied to the RegCM2 model simulation.

Finally, chapter IV expands on chapters II and III by using the 50 km simulation developed in chapter III to drive a 10 km RegCM2 simulation over Northern Utah. Only one month is simulated, but the methodology from chapter II is used to determine one of the months with the best performance of the 50 km RegCM2. The LP gridding method developed in chapter II is also used for validation of the 10 km model simulations.

Chapter IV constitutes the major thrust of this study. In it several experiments are done, and theory is developed for precipitation simulation problems unique to nested 10 km simulations. With initial experimentation of 10 km simulations we find the precipitation simulation in general to be too dry. However, over mountain up-slopes the simulation is much too wet. There are unrealistically high amounts of precipitation on the windward edges of the mountain ranges contained in the 10 km domain. After finding the problems with our model simulation, we decided to go back to basic physics and analyze the run to determine why the solution is so flawed. First we examined the literature to find which potential areas in the model physics were weak and most prone to error (especially for 10 km resolution). From the literature, several potential reasons for these problems are found. *Zhang et al.* [1988] theorize that problems with their simulations similar to ours might be due to the fact that RegCM2 is a hydrostatic model rather than non-hydrostatic. *Molinari and Dudek* [1992] blame the convective parameterization of the model, or the explicit cloud-water scheme. *Sass and Christensen* [1995] point out that the input data itself can effect the regional model solution. Simulations over the complex terrain of Japan [*Kato*, 1997] suggest that although a 10 km LAM will predict vertical velocities that are too high, the impact on precipitation is quite small. Our studies show that the convective component of the precipitation is small, and changing the parameterization does not seem to affect the simulations. We experimented with the explicit cloud water scheme to see if we could ease back precipitation near

the mountains. We found that while we could scale back the precipitation, the general nature of high precipitation over mountain up-slopes, but too low precipitation elsewhere, never changed. As a result, we were forced to turn to the analysis of the input data and the physics that happens at the interface between the 50 km simulation input and the 10 km LAM. We examine conservation of heat and water mass occuring at the interface between the two resolutions. Analysis of the water budget over the grid domain led us to examine the cloud water physics. We believe that more of the water put into clouds needs to be converted into precipitation.

With the emphasis on analysis of the 50 km simulation input to the 10 km LAM, two things become critical. First, there is a need to develop accurate methods to compare the 10 km simulations to the observed. Second, There is a need to evaluate the 50 km simulation in its own right. In order to do this we need to develop accurate methods to grid observed data to model grid-points for the simulation period. After our initial comparisons it becomes obvious that it is also important to develop error estimates for the observed data grid. The model shows much greater precipitation over the mountain up-slopes that are not seen in the observed. An important assessment at this juncture is to evaluate the uncertainty in the observed data grid and see if the observed data will allow for such a characteristic. Thus, the development of error estimation methods for the observed data grid is a critical development.

Another critical development is the analysis of the 50 km simulation in its own right. Under critical analysis we find problems with the 50 km simulation as well, although not as extreme as the problems with the 10 km simulation. However, the problems with the 50 km simulation still do not explain the behavior we observe with the 10 km precipitation.

Finally, we examine three issues in regard to the interface between the 50 km input and the 10 km LAM: temporal update frequency of the lateral boundary conditions (LBC), the way the 50 km input data were interpolated to form LBC, and the handling of the LBC within RegCM2. This analysis will be presented in chapter IV.

In summary, analysis of the problems with precipitation with the 10 km simulation (developed in chapter IV) leads us to develop the observational analysis methods of chapter II, and to do the 50 km simulation analysis in chapter III. Now we turn to chapters II, III, and IV. After we present these chapters, we will continue with the discussion and conclusions of our findings.

References

Anthes, R. A., Hsie E.-Y., Kuo Y.-H., Description of the Penn State/NCAR mesoscale model version 4 (MM4), *NCAR Tech. Note, TN-282+STR*, 66 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1987.

Arakawa A., and W. H. Shubert, Interaction of a cumulus cloud ensemble with the large scale environment. Part I, *J. Atmos. Sci.*, *31*, 674-701, 1974.

Briegleb, B. P., Delta-Eddington approximation for solar radiation in the NCAR Community Climate Model, *J. Geophys. Res.*, *97*, 7603-7612, 1992.

Corby G. A., A. A. Gilchrist, and R. L. Newson, A general circulation model of the atmosphere suitable for long period simulations, *Q. J. R. Meteorol. Soc.*, *98*, 809-832, 1972.

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, Biosphere-atmosphere transfer scheme (BATS) version 1E as coupled to the NCAR community climate model, *NCAR/TN-387+STR*, 69 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1993.

Giorgi, F., Simulation of regional climate using a limited area model nested in a general circulation model, *J. Clim.*, *3*, 941-963, 1990.

Giorgi, F., Sensitivity of simulated summertime precipitation over the western United States to different physics parameterizations, *Bull. Am. Meteorol. Soc., 119*, 2870-2888, 1991.

Giorgi, F., and G. T. Bates, The climatological skill of a regional model over complex terrain, *Mon. Weather Rev.*, *117*, 2325-2347, 1989.

Giorgi, F., G. T. Bates, and S. J. Nieman, Simulation of the arid climate of the Southern Great Basin using a regional climate model. *Bull. Am. Meteorol. Soc.*, *73*, 1807-1822, 1992.

Giorgi F., and M. R. Marinucci, Validation of a regional atmosphereic model over Europe: Sensitivity of wintertime and summertime simulations to selected physics parameterizations and lower boundary conditions, *Q. J. R. Meteorol. Soc.*, *117*, 1171-1206, 1991.

Giorgi, F., M. R. Marinucci, and G. T. Bates, Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes, *Mon. Weather Rev., 121*, 2794-2813, 1993a.

Giorgi, F., M. R. Marinucci, G. T. Bates, and G. D. Canio, Development of a second-generation regional climate model (RegCM2). Part II: Convective processes and assimilation of lateral boundary conditions, *Mon. Weather Rev., 121*, 2814-2832, 1993b.

Grell, G. A., J. Dudhia, and D. R. Stauffer, A description of the fifth generation Penn State/NCAR mesoscale model (MM5), NCAR Tech. Note, NCAR/TN-398+STR, 121pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1994.

Holtslag, A. A. M., and B. A. Boville, Local versus non local boundary-layer diffusion in a global climate model, *J. Clim.*, *6*, 1825-1842, 1992.

Holtslag, A. A. M., E. I. F. de Bruijin, and H. L. Pan, A high-resolution airmass transformation model for short-range weather forecasting, *Mon. Weather Rev.*, *118*, 1561-1575, 1990.

Hong, S.-Y., and H.-M. H. Juang, Orography blending in the lateral boundary of a regional model, *Mon. Weather Rev.*, *126*, 1714-1718, 1998.

Hostetler, S. W., and F. Giorgi, Use of a regional atmospheric model to simulate lake-atmosphere feedbacks associated with Pleistocene lakes Lahontan and Bonneville, *Clim. Dyn.*, *7*, 39-44, 1992.

Kato, T., Hydrostatic and non-hydrostatic simulations of moist convection: review and further study, *Meteorol. Atmos. Phys.*, *63*, 39-51, 1997.

Lall, U., and K. Bosworth, Nonparametric statistical inference and function estimation for hydrologic space time data, *Trends Hydrol.*, CSRI, 1-19, 1994.

Madala, R. V., Efficient time integration schemes for atmosphere and ocean models, in *Infinite-Difference Techniques for Vectorized Fluid Dynamics Calculations*, edited by D. L. Book, pp. 56-74, Springer-Verlag Publ., New York, 1981.

Matthews, D. M., F. Giorgi, and G. T. Bates, Regional model simulation of an extreme precipitation event in the Colorado river basin, in *Proceedings of the Third Symposium on Global Change*, AMS, pp. 20-23, Boston, Mass., 20-23, 1992.

Medina, J. G., Application of a simple local-scale numerical model in the study of altered climate impacts on watershed precipitation, in *Proceedings of the Second Symposium on Global Change Studies*, pp. 68-73, AMS, New Orleans, La, 1991.

Molinari, J., and M. Dudek, Parameterization of convective precipitation in mesoscale numerical models: A critical review, *Mon. Weather Rev.*, *120*, 326-344, 1992.

Sass, B. H., and J. H. Christensen, A simple framework for testing the quality of atmospheric limited-area models, *Mon. Weather Rev.*, *123*, 444-459, 1995.

Zhang, D., E. Hsie, and M. W. Moncrieff, A comparison of explicit and implicit predictions of convective and stratiform precipitating weather systems with a meso-Beta-scale numerical model, *Q. J. R. Meteorol. Soc.*,114, 479 31-60, 1988.

CHAPTER II: DOWNSCALING GCM SCALE OUTPUT TO MOUNTAINOUS REGIONS USING A REGIONAL MODEL: PART I: COMPARISON METHODOLOGY

Abstract

Many groups use regional climate model data to drive ecological, hydrological, and surface interaction models. However, direct long-term comparisons of regional model skill are limited, especially in mountainous terrain. In this chapter we develop the methodology to compare model data from a regional model to surface observed datasets. To compare these, spatial scale and elevation differences between regional model output and station data must be resolved. We demonstrate a robust method to compare a model simulation with observed data. Our gridding technique uses local polynomial regression in latitude, longitude, and elevation to estimate precipitation and temperatures. We compare this method with a less rigorous elevation adjustment scheme. The third method is the average over the observed stations in the region. With all methods we develop a statistical estimate of the observed error for the resulting grid. Our comparison statistics examine long-term signal, daily forecasting ability, and long-term frequencies.

II.1 INTRODUCTION

Impact studies of potential greenhouse warming on regional climate have caused an explosion of effort into downscaling global climate model (GCM) output. Downscaling efforts include using regional climate models for a physically based approach, to using empirical methods based on observational data [*Giorgi and Mearns,* 1991]. In order to use regional climate models for this effort, it is important to verify model results by comparing to the observed climate. The primary emphasis of this chapter is in presenting a comparision methodology of regional climate model output to observed data. A secondary emphasis is in presenting methods for creating datasets for input into impact models based on observational data. The datasets required to substantiate the regional climate model (a gridded dataset) are the same as the type of datasets required for use by impact models.

Two critical data-types for regional climate model comparisons are precipitation and temperature. In mountainous regions like the western US, the high elevations experience much higher precipitation rates and store the water longer due to the occurrence of snow and lower evaporation rates. In this area over 70% of the runoffs occur from less than 25% of the total surface area. The lower elevation regions, which are typically desert, receive less than 20 cm per year, while the higher elevations of these regions often receive more than 125 cm per year. A particular test both for gridding observational data and regional climate models is to properly distribute rain shadow affects [*Cotton and Anthes*, 1989; *Sumner*, 1990]. Upstream slopes experience much higher and longer precipitation events than those on the down-slope side. Another challenge is that existing observational sites are not evenly distributed throughout the elevation range of the surface. Until a few years ago, almost all data collection occurred at sites several hundred meters below most model surfaces [*Giorgi,* 1990]. This problem has been eased by the placement of new observation networks in the late 1980s to determine wintertime snow accumulations. These stations are at higher elevations but they tend to be sited in the broader, snow-catching regions and undersample the higher elevations. There are also far fewer of these stations than the National Oceanic and

Atmospheric Administration Cooperative (NOAA COOP) data stations, which tend to be at lower elevations.

Accurate comparisons of regional model outputs in mountain areas with observed data require major effort [*Giorgi and Bates,* 1989; *Giorgi,* 1990; *Giorgi et al.,* 1993; *Barros and Lettenmaier,* 1993; *Leung and Ghan,* 1995]. One study published by *Giorgi et al.* [1993] compared RegCM output and 390 US NOAA COOP stations distributed across the western US. This analysis compared the model output to observed data by averaging the output over a region, and then averaging the observed station data over the same region. These and other elevation-averaged comparisons [*Giorgi et al.,* 1994; *Luthi et al.,* 1996; *Jacob et al.,* 1997] leave concern about how the hydrologically important mountain regions are being treated. What the *Giorgi et al.* [1994], *Luthi et al.* [1996], and *Jacob et al.* [1997] studies did was to ignore the elevation of a site and simply average all observed stations within a region – calling the result the regional response. However, since this method significantly undersamples the higher elevations – known to have significantly higher amounts of precipitation – the end result undersamples the true regional average. This places under suspicion any conclusions about the model comparison when the observed dataset is known to be too dry. Even this would be acceptable if the observed datasets had error estimates. Unfortunately, this is not the case, only leaving the reader to wonder about the conclusions of the study. The *Giorgi et al.* [1993] analysis showed a 600 m difference in height between model and observed stations.

In this chapter we demonstrate comparison methods to compensate for the elevation differences between model and observed. Assuming the empirically observed average temperature lapse rate of -6.5 C/km, a 600 m rise in elevation accounts for a 4 degrees Celsius bias in temperature. For our comparison we will use two documented station data gridding techniques [*Lall and Bosworth,* 1994; *Briggs*, 1974], which we will also compare to a straight average of all station data in a given region as done in the *Giorgi et al.* [1993] analysis. Our comparison methodology ensures that the same quantities are being compared from model to observed data. Model area averages are compared to observed area averages, and observed temperatures taken at screen height are compared to model temperatures adjusted to screen height. We provide observed error estimates because they are needed to determine the quality of the data. And we also use a wider variety of statistics to judge not only the long-term responses, but also the frequency response, and spatial patterns.

II.2 DISCUSSION

In order to have a valid comparison of observational data with modeled data, it is important to ensure that all data-types represent the same physical quantities. First this requires either converting the area-averaged data of the model to point data, or gridding the observed data onto the same grid as the model. To grid, we take the observed data at scattered locations and use statistical methods to come up with an estimate of the model grid locations. Because we do not want to introduce additional uncertainty in the model data, we choose the gridding method. Second, this requires that both the observed and model temperatures be taken at screen height. Screen height is the height at which temperature observations are taken and is typically in the range of 1.2 – 2 m. This means using the RegCM2 model formulation for the surface boundary layer to adjust the lowest model layer temperature (around 20 m) to screen height.

II.2.1 Observed Datasets

The observed datasets used in this comparison were developed by combining quality controlled versions of data from NOAA COOP and the US Department of Agriculture National Water and Climate Center Snow Telemetry network (SNOTEL). The COOP network has been in place and stable for over half a century. The SNOTEL network was still developing in the mid-1980s and the number of stations reporting temperature data changed rapidly after 1987.

II.2.2 Adjustments to Model Data

We must ensure that both datasets represent the same physical quantity in order to provide a fair comparison between observed and modeled data. Primarily this requires gridding the observed point dataset so that both datasets represent areal averages of the same resolution. There are a few adjustments required to the model dataset to make it consistent with the observed data. Let us examine these adjustments.

The data-types we compare to the model data are observed daily maximum temperature, minimum temperature, mean temperature, and precipitation. Observed temperatures are taken at screen height. However, the model data are stored on the lowest model layer (roughly 20 m off the surface). To adjust the modeled temperatures to screen height, we use the hourly data from the lowest model layer and a formulation consistent with the surface boundary layer physics in RegCM2 (the BATS1E formulation) [*Dickinson et al.,* 1993]. These same expressions are used inside the regional model to model the interaction of the surface with the atmosphere. In order to provide physics consistent with the methodology used in the regional model, we use the same formalisms as used in the model. Even so, there are several reservations that we have with the following methodology. We will discuss the reservations after laying out the methodology.

Equation 1 shows how the model physics is used to find the potential temperature at screen height based on lowest model layer temperature. Equation 2 derives the potential temperature of the lowest model layer.

$$\theta_{sh} = \theta_m - \theta^* \psi(RI_b) \log[Z_m/Z_{sh}] / (C_p \rho K u^*) \quad (1)$$

$$\theta_m = T_m * (P_m/P_s)^{\alpha} \quad (2)$$

where θ_* is the sensible heat flux (see equation 6); Z_m is the height of the lowest model layer; Z_{sh} is screen height (1.5 m); θ_m is the potential temperature of the lowest model layer; θ_{sh} is the potential temperature of the resultant height; ρ is the air density; C_p is the heat capacity for air at constant pressure (1004 J kg^{-1} K^{-1}); and K is the Von- Karmen constant (0.378). P_m is the pressure of the model layer; T_m is the temperature of the model layer; P_s is the surface pressure; $\alpha = -R/C_p$, where R is the gas constant for dry air, (287 J kg^{-1} K^{-1}) ψ depends on stability as given in equations 3 and 4. For neutral conditions ψ is one; and in RegCM2 (BATS1E), it is dependent on the Bulk Richardson's number (RI_b) as follows.

For stable conditions ($RI_b > 0$)

$$\Psi = \sqrt{(1 + 11.5 * RI_b)} \quad (3)$$

For this case $RI_b = \min(RI_b, 3/11.5)$

For unstable conditions ($RI_b < 0$)

$$\Psi = 1 / \sqrt{(1 + 24.5*K*\sqrt{(RI_b)/\log(Z_m/ Z_{sh})}} \quad (4)$$

The Bulk Richardson's number is a common stability term

$$\mathbf{RI_b = (gZ_m/\theta_v)*(\Delta\theta_v/|U_m|^2)} \quad \mathbf{(5)}$$

where g is gravitational acceleration; θ_v is the virtual potential temperature of the lowest model layer; $\Delta\theta_v$ is the difference in virtual potential temperature from surface to lowest model layer; and U_m is the wind speed at the lowest model layer (plus a "convective" component 0.1).

The sensible heat flux is derived from the potential temperature difference between surface (see equation 9) and lowest model layer as follows.

$$\mathbf{\theta_* = C_p C_D \rho U_m T_{skin} - \theta_m)} \quad \mathbf{(6)}$$

$$C_D = C_{DN}\, \psi^2 \quad (7)$$

where we limit C_D such that it can not be less than $\max(C_{DN}/4, 6x10^{-4})$.

$$C_{DN} = K^2(\sigma_{veg}/\log(Z_{sh}/Z_{veg}) + \sigma_{soil}/\log(Z_{sh}/Z_{soil}) + (\sigma_{snow}/\log(Z_{sh}/Z_{snow}))) \quad (8)$$

where $Z_{veg}, Z_{soil}, Z_{snow}$ are the roughness heights for vegetation, soil, and snow, and σ_{veg}, σ_{soil}, σ_{snow} are the fractional areal-coverage of vegetated, soil, and snow-covered surfaces. Because of this, θ_{sh} is bound between the range of θ and T_{skin}. In RegCM2 (BATS-1E), the skin temperature is defined as the weighted-average of the soil and canopy temperature by the following equation, where the weight is the fraction of area covered by vegetation:

$$\mathbf{T^4_{skin} = \sigma_{veg} T^4_{veg} + (1 - \sigma_{veg}) T^4_{soil}} \quad \mathbf{(9)}$$

where σ_{veg} is the fractional vegetative cover; T_{veg} is foliage temperature; and T_{soil} is the soil (or snow) surface temperature.

The data are stored at hourly intervals and T_{max} and T_{min} are the respective maximum and minimum hourly values for the simulated day. T_{ave} is the simple average of T_{max} and T_{min} for that day. This offline procedure adjusts for the height difference in a way consistent with the surface physics contained in RegCM2. This method makes a significant difference in the reported temperature.

At the beginning of the section some reservations with this methodology were mentioned. The first concern is the fact that similarity theory is developed by establishing empirical relationships in flat homogeneous terrain under steady-state conditions and sufficiently large averaging periods, and yet we apply these relationships over complex terrain. BATS1E assumes a flat surface, and while the large 50 km grid-squares have slope, they are large enough that the slopes are very small. BATS1E also assumes a single vegetation type within a model grid-square; however, it allows for three types of surfaces within a grid-square: vegetated, bare, and snow-covered. Under some conditions one of these surfaces may represent a very small fraction of the grid-square and, as

such, deviate from the homogeneous surface type. The model is solving for the mean state and not the perturbed state and as such satisfies the steady-state condition. It is common practice in boundary layer theory to separate the mean from the perturbed state and to solve for the mean state, which is what we are doing here. These issues become larger at finer and finer resolutions, but are well in check for resolutions of 10 or 50 km. Studies that compare different surface models used in global or regional climate models show significant variation between the different models. But, as all the models rely on similarity theory, it suggests that differences in the physics of other areas are more important in explaining these differences.

In the idealized model world the relationships are reasonable but, in the real world these relationships do not hold. Fluxes at the sub-grid-scale are much more complex than the model can simulate. This is one reason that analysis needs to be done over fairly large areas and long periods. The model simply cannot hope to capture the fine detail for a specific time. Coherent sub-grid fluxes are also a likely source of some of our model error when compared to observations.

Another area of concern is the expression for the drag coefficient. The dependence on stability was derived from an expression suggested by *Deardorff* [1968]. This expression was based on a study over ocean and modified for use over land. It was significantly different from other widely used and accepted expressions based on recent experiments over land – the Businger Dyer relationships [*Dyer,* 1974]. For neutral conditions the expressions are the same, but the values diverge. The expressions also give a limit on the range of RI_B that the expressions are valid for, but the model frequently goes outside that range. There are various limits on quantities so that values remain in check, but the formula is outside its intended range of applicability.

For precipitation, we total both the convective precipitation and the grid-scale resolvable precipitation. We also verify that it is greater than zero. If it is not, we set it to zero, as under rare conditions RegCM2 produces precipitation of slightly negative value. We were not able to determine why RegCM2 produced negative precipitation, but it only happened on rare occasions and had no impact on the overall quality of the model. Because we use gridding techniques to bring observed precipitation and temperatures to the model grid position and elevation, both datasets are directly comparable.

II.3 COMPARISON METHODOLOGY

To compare model data to observed data we need to have a common reference. This requires gridding the observed data onto the same grid as the model. Several techniques are available to grid observed station data. Rather than add additional uncertainty to the modeled data, we choose to compare the two techniques (gridded observed data and regional model output) on the same model surface grid-points. Since the observed gridding techniques reconcile elevational differences, the comparison on this surface is valid. Two types of averaging can be applied -- an area average of all the grid-points for any given time, and time averages. All of our plots are area averages of all the data-points within the region. There are only three exceptions to this rule, which we will discuss later. Most of the plots also show time averages (monthly, annual, or period) of the area averaged data. Climatology plots are the period averages for a given month.

II.3.1 Comparison Regions

We will examine our gridding methods over 11 small 250 x 250-km square regions spread throughout the western US. These regions are the same regions that we use to compare to the model in chapter III. See Figure 1 for a digital elevation model (DEM) of the western US outlining each region and presenting its precipitation climatology. Table 1 lists the abbreviations and numbering we use to identify the regions. Most of the examples we show will be for one of the regions, the northern Utah region.

II.3.2 Observed Data Gridding Methods

In order to verify our new gridding technique we will compare two methods. The first method (average estimate) calculates data for a new elevation based on a linear regression with altitude of the data from all stations near the gridded point. The second more rigorous approach calculates data for the grid-point based on a polynomial regression involving elevation and location (latitude and longitude). These methods are detailed below. Both methods use the same data input and hence draw from the same neighborhood. They also use the same error checking procedure outlined below.

At 50 km resolution the observation network is reasonably dense. Each comparison region has around 100 observation stations for the 25 50 x 50-km grid-points we use. See Figures 2 and 3 for a plot of the station locations and the grid-points used for each region.

II.3.3 Observed Data Error Checking

Automatic station measurements sometimes contain obviously bad data. A basic error-checking scheme is applied to all of the observed data input. The error-checking scheme is as follows:

1) For the set of y_i observed data-points over the entire region being examined at the time in question, find the median of the nonduplicated values (M_y). The nonduplicated median is found by taking the median of the series where only one value from any duplicated value is used.

2) Find the absolute deviations from this median ($d_i = |M_y - y_i|$).

3) Find the median nonduplicated deviation (M_d). This is done by sorting the data and throwing away any extra data-points that duplicate the values of other data-points. After duplicate data is thrown away the median is found (M_d).

4) For each data-point calculate the ratio of the absolute deviation of a point to the median deviation ($r_i = |M_y - y_i|/M_d$). For daily precipitation, a point is assumed to be a bad data value only if:

$$r_i > 7.5 \text{ and } ((y_i - M_y) > 25 \text{ mm}). \qquad (10)$$

This amounts to throwing out data that exceed 7.5 times the standard deviation. Assuming a normal distribution, this corresponds to a probability less than 0.001% that this could happen by random chance. By standard statistics this seems excessively low. However, since the precipitation data are not normally distributed and the sample size is relatively small, a large threshold is needed to ensure most data are kept. These values are found by applying the error checking to region 8 and ensuring that obviously bad data are thrown out, but reasonable data are kept. We discovered

that even such a large ratio could not always prevent reasonable data from being rejected. In some cases points with a small deviation from the median are rejected simply because r_i was large. In order to only reject points with a large deviation, we add a rejection threshold so that only points that deviate from the median by more than 25 mm are rejected. Thus, precipitation values between zero and M_y are still considered to be valid, and values greater than 7.5 are valid if the absolute deviation is relatively small.

Temperature points are assumed to be bad data if:

$$r_i > 7 \text{ and } |M_y - y_i| > 10 \text{ C.} \quad (11)$$

In addition, the data are range trimmed (deleting data outside reasonable bounds for the given data-type). A daily precipitation point is assumed to be bad if it falls outside the range (0 to 150 mm), and temperature points are assumed to be bad if they fall outside the range (-60 to +60 degrees Celsius). Nonduplicated medians are required for precipitation because it is common to have several points with no precipitation and only one or two points with precipitation. Without using the nonduplicated median, the dry points significantly bias the statistics and cause even reasonable precipitation amounts to be thrown out. For consistency we use the same procedure for temperatures.

II.3.4 Neighborhood Definition

At least 7 points (n_{min}) from the surrounding 150 km are used to calculate the estimate of the observed area average at each model grid-point. These points are chosen as follows. All of the points within a block radius in latitude and longitude of (r = H*1.5/(110 km/1 degree)) of the center point are selected. H is the model resolution of 50 km, and 110 km is the approximate arc-distance covered on the globe over one degree. After points are chosen, bad data-points are dropped so that at least n_{min} points can be used for estimation. If there are not enough points, the radius is increased 50% and tested again. If it still does not include n_{min} points, it is increased another 50%. The program allows a maximum of four scale increases, but the need for increasing the radius is uncommon. On rare occasions one increase is necessary, but four increases are never required. Note that the neighborhood needs to be chosen for each time. This is because the observed stations are under operation at different time intervals, and bad data-points have to be filtered out for each time.

In Figures 2 and 3 we show a plot of the station locations as well as the model grid-point locations. We see that in general there are more stations than model grid-points. Most areas are covered but there are still some regions of sparse coverage. However, this coverage is of high enough density that we can have confidence in the gridded dataset.

II.3.5 Observed Data Gridding Techniques

Two observed data gridding techniques are applied in this comparison. The local polynomial (LP) approach provides a sophisticated comparison to the model data. The average estimate (AE) approach is intended to provide a rather simple comparison. The final comparison method is the region average method (AVG), the simple average of all the observational data stations in the region.

II.3.6 Local Polynomial Regression

A common approach to gridding climate data in mountainous terrain is to orographically adjust the temperature assuming a given temperature lapse rate (typically assuming a 6.5 degrees Celsius drop in temperature per 1000 m rise in elevation). Adjusting precipitation requires a more sophisticated relationship. Even the orographic relationship of temperature can significantly deviate from the average lapse rate such as in temperature inversions. Therefore, statistically finding the relationship of temperature with elevation is more useful than assuming a given lapse rate. This is especially true when examining daily data in a highly localized area.

The LP approach adjusts the observed data elements to the required grid-point location using a polynomial regression over position and elevation. Because the data are weighted by proximity to the required grid-point and elevation, the error estimate is smaller than one based on a straight linear combination of stations.

The methodology for using LP regression for spatial estimation of hydroclimatic variables was developed by *Lall and Bosworth* [1994]. Consider a general regression model given as:

$$\mathbf{y_i = f(x_i) + e_i} \qquad \mathbf{i=1,...,n} \qquad \mathbf{(12)}$$

where x is an M_r dimensional-vector (three in our case) of location coordinates; y is the response variable (precipitation or temperature); f(x) represents the underlying functional relationship between y and x; e_i is noise or measurement error that may or may not depend on x_i; and n is the number of observations.

An approach for the pointwise estimation of the unknown function f(x) from data, based on a local Taylor series expansion of f(x) at the point of estimate x, was proposed by *Macauley* [1931]. *Cleveland* [1979] pioneered this idea into a statistical methodology for local approximation of functions from data [*Cleveland and Devlin,* 1988; *Cleveland et al.,* 1988]. A recent monograph by *Wand and Jones* [1995] provided a general background on the methods.

The strategy is to choose a certain number, k, of nearest neighbors (in terms of Euclidean distance) of the estimation point x. Then we form the estimate f(x) through a locally weighted, polynomial regression over the (x, y) data that lie in the neighborhood. When making point estimates the number of neighbors is chosen to minimize the error. In our case we want to estimate areal averages, so we take all of the points within the current grid-box, and the grid-boxes one grid-box away in any direction from the current one (using the neighborhood definition derived in section 3.2.2). Let D be the set of x_i that are within the k nearest neighborhood defined in section 3.2.2. By defining D in this manner and restricting the estimate to a linear regression, the end result is an areal average rather than a point estimate. Consider the general regression model described in (1). The sampling locations (x_i) are usually not regularly spaced. We assume the e_i associated with D are uncorrelated, random variables with a zero mean. We also assume that the e_i are approximately evenly distributed throughout D.

Then, the locally weighted polynomial regression at each point of estimate x_l, l =1..np, given a (n x M_r) data matrix X and a (n x 1) response vector y, is obtained through the solution of the weighted least squares problem:

$$\underset{\beta_l}{\underline{\text{Min}}} \; (y_l - Z_l\beta_l)^T \, W_l \, (y_l - Z_l\beta_l) \qquad (13)$$

where the subscript l recognizes that the associated element is connected with the point of estimate x_l; β_l are estimates of the coefficients of the terms in the basis defined by Z_l; Z_l is a matrix formed by augmenting X, with columns that represent the polynomial expansion of X to degree p (including cross product terms if desired). In our case p is 1, linear. W_l is a k x k diagonal weight matrix with elements

$$\mathbf{w_{ii,l} = K(u_{i,l})/\sum K(u_{j,l})} \quad \textbf{(14)}$$

$$K(u)=15/16(1-u^2)^2 \quad (15)$$

where $u_{i,l} = d_{i,l}/d_{k,l}$; $d_{i,l}$ is the distance from x_l to x_i using an appropriate metric; and K(u) is a weight function. For the distance metric we used the sum of the distances in each coordinate (each coordinate is normalized to [0,1]). We keep the regression linear so that the end result represents the area average of the entire grid-square.

We have implemented a bisquare kernel weight as shown in equation 14 above. This is recommended by *Scott* [1992] because of its smoothness properties. Let D_l be the k nearest neighborhood of x_l (as defined in section 3.2.2). Then the matrices y_l and Z_l are defined over D_l. Singular value decomposition (SVD), with algorithms from *Press* [1989], is used to solve the linear estimation problem resulting from equation 12.

The coefficients β_l are obtained as:

$$\beta_l = (\mathbf{Z}_l^T \mathbf{W}_l \mathbf{Z}_l)^{-1} \mathbf{Z}_l^T \mathbf{W}_l \mathbf{y}_l \quad (16)$$

The resulting estimate of $f(x_l)$ is then:

$$\mathbf{f(x_l) = z_l \beta_l} \quad \textbf{(17)}$$

where z_l is the d'1 vector formed by augmenting x_l with polynomial terms to order p, and retaining the terms for which β_j are found to be significantly different from 0.

Lall and Bosworth [1994] introduced the use of a local GCV score that uses data directly from the local regression at the point of estimate. The local GCV score provides a measure of the variance of the predictive error at each point of estimate. In this case, the errors $e_{i,l}$ are the residues of the model fitted over D_l, and W_i is the corresponding weight matrix. The local GCV score is then given as:

$$LGCV_l(\hat{f})=\mathbf{e}_l^T\mathbf{W}_l\mathbf{e}_l/((k-d')/k)^2 \quad \textbf{(18)}$$

where d' is the number of coefficients fit in the local regression. An estimate of the standard error of prediction of the local regression is thus obtained as $\sqrt{(LGCV)}$.

II.3.7 Average Estimate

As a comparison to the LP approach, we apply a baseline approach to the gridding process. In this baseline approach, we compute the observed estimate at the model grid-point using an AE technique. In this method, a least squares linear regression with elevation of the observed data

within the model grid-point neighborhood is run for each data element (maximum temperature, precipitation, etc.). An estimate of the data element value at each model grid-point is computed for each observation station in the grid-point neighborhood (as defined in section 3.2.2). The grid-point average and an error estimate are then calculated. The error estimate is chosen using the standard deviation about the resulting average:

$$(\sigma_{dev} = \sqrt{(\Sigma((y_i^* - y_l')^2) /k)) \quad (19)$$

where y_i^* are the data-points adjusted to the predicted elevation using the slope of the linear regression, multiplied by the elevation difference, and y_l' are the estimates at the model grid centers. Note that for this case the definition of the LGCV (used for LP error estimates see section 3.4) is almost identical to the variance (σ_{dev}^2), because the weights are uniform for a straight average. The region average is then calculated using the arithmetic mean of all estimates in the region of interest.

II.3.8 Regional Average Method

Next we examine the comparison of our gridding methods (LP and AE) to region averaged studies (AVG). The total error is the sum of the data error plus the method bias (the error due to the difference in elevation and position). To approximate the data error we use the standard deviation of the data in the region. And to approximate the method bias we use the difference between the AE method and the region average. The data error is larger than the errors for LP and AE since it is applied over a larger area, and does not compensate for the bias due to differences in elevation. The addition of method bias due to elevation and position differences makes the total error in this method unacceptably large. In Table 2 we show the average the elevations of the model grid-points within each region. We also show the average elevation of the actual observed stations within the region. Since observational weather stations tend to cluster in population centers, they also tend to cluster at lower elevations. Few exist in mountains and even fewer near mountain peaks. Hence, the average station elevation tends to be much lower than the average elevation of the region. Since the model topography comes from DEM data, it tends to more closely reflect the true average elevation of the region. Thus, when in Table 2 we also show the difference between model and observed average elevations, we see that the model is almost always at a significantly higher elevation than the observed. That average difference is quite large (417 m), with the smallest difference being 209 m, and the largest, over 1000 m. Differences in elevation this large will also translate into errors in average temperature and precipitation, if the average of the station data are used to derive them. Assuming a temperature lapse rate of 6.5 C/1000 m, this translates to an error of 2.7 degrees Celsius. An approximation of the method bias for precipitation is harder to judge, but it is obvious that it would be significant simply because of the bias due to elevation difference. The method bias due to horizontal shifts in position is small, but can be significant if the shift goes into or out of a rain shadow region.

To specify values to use for the data error and method bias, we use the same error estimate as for the AE method. The data error for the AVG method is the standard deviation of the data-points. The AE method and the AVG method are quite similar. Both are simple averages. The AE method, however, compensates for the method bias by running the regression in elevation. The AE method also calculates for each grid-point before calculating the region average. Thus, the method bias for the AVG method is simply the difference between the AVG and AE methods. And, the

total error is the sum of the standard deviation and the difference between the AE and AVG methods. It is important to realize that because the elevation differences are not compensated for, there is a much wider variation in data values in the AVG method than the AE. Thus the standard deviation is much larger because of this variation. And because the AE method uses the local data to calculate each grid-point, the variation is also reduced because it examines only the data nearby to the grid-point.

II.3.9 Comparison Statistics

We develop a suite of comparison statistics oriented from the perspective of how well a model would compare to the observed. Table 3 lists a summary of the comparison statistics we will use. In chapter III we use these statistics to examine the modeling ability of the RegCM2 model. In this chapter we demonstrate the statistics with our different gridding methods (AE, LP, and AVG). The methods we develop can be applied over very small regions up to the entire modeling surface. The goal is an examination of the simulation for the long-term climate of the model. We are not necessarily looking at how well the model reproduced the observed series. However, we examine the whole period to better understand in what aspects the simulation is good and in what ways it is weak. We also compare the model bias to the observed error, compare model bias to observed standard deviation, and compare the model results to a forecast model that uses observed climatology. The long-term climatology comparisons provide the long-term biases of the model, and the threat score provides the short-term forecasting ability. These statistics give us a comparison to judge the model bias in a quantitative way.

II.3.10 Climatology

Our first statistical comparison between the methods is "climatology." This is the average of each month for all the years over the simulation period. We also examine both observed and model standard deviation calculated for each month. Since short model simulations can reproduce (or conversely fail to reproduce) the observed for the wrong reasons, it is useful to examine a model's response over a longer time-period.

The next statistic we want to estimate is the model bias, or deviation of the model from observed. A good approximation of the long-term model bias is simply the difference between the model and observed climatologies.

Since a climate model's bias may be large, the next question to ask would be if the model tracks the deviations in the observed sequences. Monthly anomalies are the deviation from the climatology for each month. Anomaly correlation is the report of the correlation coefficient between the observed and modeled climatology anomaly.

$$\mathbf{Corel}[(o - o_{clim}), (m - m_{clim})] \qquad (20)$$

where o = observed monthly data and o_{clim} refers to the average of the given month over the entire simulation period. Similarly, m = model monthly data and m_{clim} refers to the average of the given month over the entire simulation period. A simple significance test is to examine two uncorrelated Gaussian distributions. For this case the 95% significance level of the correlation is $2/\sqrt{(n)}$, where n is the number of observations.

The skill is a comparison of the model element to a simulation using the observed climatology (i.e., does the model perform better or worse than simply using climatology as the prediction). To compensate for possible long-term biases in the model, we subtract the long-term difference between the model and observed climatologies. This way the skill examines the model simulation in terms of matching the response of the observed inter-annual variability. Examining the climatologies already provides insight into the long-term model biases.

$$\mathbf{Skill = 1 - (|o - m - (o_{clim} - m_{clim})|/median(|o - o_{clim}|))} \quad \textbf{(21)}$$

II.3.11 Threat Score, Bias Score, and Monte-Carlo Threat Score

The threat score and bias scores for various precipitation thresholds are statistics commonly used in forecasting. *Giorgi et al.* [1993] also successfully used this statistic in analysis of RegCM model simulations. Similarly we use the long-term average of these statistics to test the model's predictive skill throughout the simulation. The threat score indicates how well the amount and placement of precipitation is modeled or forecast. This is defined for several precipitation intensity thresholds. Each threshold is examined to see how well the model forecasts the pattern for precipitation that exceeds that threshold amount. By looking at small to large thresholds, we see how well the model forecasts light to heavy precipitation events. The definition of threat score is:

$$TS = F/(M+O-F) \quad (22)$$

where M is the number of model points where the precipitation is greater than a given threshold; O is the number of points where the observed is greater than the threshold; and F is the number of points where both observed and modeled data are greater than the threshold. Because our observed data has associated error, we need to measure the threat score in such a way that any modeled precipitation within those error bounds will give the same result. Our definitions, then, become:

$$\mathbf{err_i = observed\ error} \quad \textbf{(23)}$$

$$th = \text{precipitation threshold} \quad (24)$$

$$M = \#\text{ of points such that: } m > \max(0, th\text{-}err_i) \quad (25)$$

$$O = \#\text{ of points such that: } o > \max(0, th\text{-}err_i) \quad (26)$$

$$F = \#\text{ of points such that: } (m > \max(0, th\text{-}err_i)) \text{ AND } (o > \max(0, th\text{-}err_i)) \quad (27)$$

The threat score evaluates the forecast for a given area and precipitation threshold. The area can be as small as two points and up to the size of the grid domain. By applying the threat score over the entire model domain, we can see the performance for the entire area. But, to understand how performance varies within the grid, we apply it to localized regions. Although higher numbers of points result in better statistics, smaller areas have the advantage of localizing exactly how the forecast performs in those areas.

Bias score, like the threat score, uses precipitation thresholds. It is the ratio between the number of points exceeding the threshold for the model, and the number of points exceeding the threshold for the observed. The bias score uses the same definitions used for the threat score as defined in

equations 23 through 27. The threat score examines how well the precipitation matches the pattern for a given threshold. However, there are three reasons for an imperfect threat score. Either the model has too many points exceeding the threshold, too few points, or they are simply in the wrong locations. The bias score is able to tell us if there are too many or to few grid-points that exceed the threshold.

$$\mathbf{BS = M / O} \qquad \mathbf{(28)}$$

Thus, the bias score for a perfect model would be 1.0. If the bias score is less than 1, the model has too few points above the threshold; if greater than 1, it has too many points above the threshold. We can more or less interpret the bias score as the fraction of the observed precipitation area that the model has precipitation for a given observed storm. Thus, a value of 1.5 indicates that the model has precipitation over a area 50% larger than observed. Likewise, a value of 0.5 indicates that the model only precipitates over an area half as large as observed. The different thresholds then correspond to the different levels of precipitation. A BS of 0.1 for a threshold of 10 mm indicates that the model only has 10% of the observed precipitation area of 10 mm or above. Higher thresholds typically result in lower BS values since the model not only needs to trigger precipitation at the right time (and near the right location), it also needs to trigger the right quantity.

A model's threat score can be high for low precipitation thresholds. In order to determine the possibility that random chance would result in high values, we built a statistic that we call the Monte-Carlo threat score (TS_{mc}), that is similar to the threat score, but subtracting out the signal due to random chance. Thus:

$$TS_{mc} = TS - (\sum TS\ (o|o_{rand})\ ) / N \qquad (29)$$

where we set N = 1000; TS is the previously calculated threat score between modeled and observed and $TS(o|o_{rand})$ is the threat score calculated using random observed data. The idea is that we want to compare the model's threat score with the lowest order stochastic model. This way scores near one indicate real forecasting ability, while scores near zero indicate the model's forecast is no better than a noise pattern. The resulting threat score for the random case is calculated and averaged over a large sample size (N=1000). The idea behind the stochastic model is to pick random data values for each grid-point. Thus the stochastic model uses realistic data values, but the precipitation patterns will be a poor forecast of the actual observed pattern. By averaging this over randomly generated "stochastic forecasts" we estimate the threat score for a random signal. To pick realistic values for each grid-point we arbitrarily select a different day in the same month, and use the associated value as the value for the stochastic model. We prevent using the same day as we are forecasting to give realistic but not perfect signals. We also apply a different random choice of day for each grid-point, which ensures the resulting pattern is uncorrelated with actual observed patterns. Use of the TS_{mc} then indicates for which threshold levels the model has no forecasting ability.

II.3.12 Precipitation Frequency

The threat score provides the short-term forecasting ability and the long-term climatology comparisons provide the long-term biases of the model. We also examine how the frequency of

precipitation events compares to the model. Looking to see if the duration and the intensity of those events are similar to the observed is also important. To investigate these issues, we estimate the probability density function (PDF) for dry-spell (number of days between precipitation events) and wet-spell (number of days with continuous precipitation events) lengths. To examine the frequency for different intensities of precipitation events, we estimate the probability mass function of intensity (PMF). Another method commonly used to do this is to use histograms (where we bin observed spell lengths or intensities into different classes). For example, *Giorgi et al.* [1993] used histograms for intensity analysis of several of their runs. However, *Silverman* [1986] reported problems with their use, primarily because they are sensitive to class width and origin of each class. Histograms do not clearly show the significance of variations in the frequency. If class sizes are small, a space between observed points may be randomly selected. The resulting signal tends to be very noisy or primitive with too few classes. Using continuous functions, we gain a better understanding of the actual signal and the important variations in it. Our estimates of PDF and PMF use the kernel density technique of *Lall et al.* [1996] and *Rajagopalan et al.* [1996, 1997]. The method involves centering blocks at each observed point and using a function that tapers off away from that point, rather than a rectangle. These kernel functions are applied over each point and then summed to get the resultant PDF. For points closely packed around a given interval, a higher PDF is produced at that point. The final result gives smoothed functions that are better suited to displaying how precipitation intensity and frequency vary over a given interval. The kernel functions we use are normal functions as follows:

$$K(t) = (2\pi)^{-1/2} e^{-t/2} \tag{30}$$

where $t = ((f - f_i) / h)^2$; f = wet/dry spell or precipitation intensity value (quantity that we wish to determine the frequency distribution of); f_i = actual observed such value; and h = the bandwidth chosen with the method of *Sheather and Jones* [1991].

II.4 RESULTS

Now we use some of the statistics we have generated to compare the gridding methods. Because the AE and LP methods are so close many of the statistics are not useful for comparison. For example, the anomaly correlations between LP and AE are over .98 for all regions and data-types. What we display are the error estimates, and the differences in climatology. Also we only display T_{ave} and precipitation. The maximum and minimum temperatures display the same biases that exist in the average temperature. When comparing to a model, maximum and minimum temperatures are important to analyze since the model may have a bias in the daily temperature extremes. In chapter III we use the LP method to compare model to observed over 11 different regions in the western US.

II.4.1 Observed Data Error Checking Method

First we want to examine the data error checking of the AVG method and its impact on the resultant answer and error estimates. As a simple check, we examine the impact of screening on the averages for all precipitation data in our 11 selected regions. Table 4 shows the maximum and mean differences between the screened and nonscreened data for each of our sampled regions. The amount of data screened on average is very small (about 2%). The average difference between unscreened and screened data is small. The highest percentage of data screened is also minimal,

31% over all regions and times. However, the maximum differences between screened and unscreened averages are very large. Thus the screening is effective in screening out bad data-points that would significantly alter the results. The screening is invoked only a small percentage of the time (8 to 31% depending on the region). Thus our algorithm is effective in screening out truly bad data, but in most cases it is not even invoked. This is exactly what we hoped the methodology would do. Also note that only half of the stations shown in Figures 2 and 3 are used in the grid-point estimation.

II.4.2 Difference in Error Estimates

The first and main evaluation of the difference between the observed gridding techniques is to compare the error estimates between the methods. Figure 4 and Figure 5 show the error estimates for the LP, AE, and AVG methods for each data-type and region. The LP method is significantly better than the AE method for both temperatures and precipitation. For the AVG method, both the base error and the total error due to its bias is shown. The AVG method has such large errors for these mountainous regions that the method becomes unusable. In all cases, the error estimates for LP and AE are much less than the AVG method, with the error estimates for LP being the smallest. However, even the 35 mm error that the LP gives is substantial. Although less than the predicted values, it represents the majority of the observed signal. As such the value is too high to have confidence in the data created. It turns out the reason for the high error values for precipitation is the fact that we took daily precipitation and totaled it to get the monthly values. Hence, the daily errors are totaled into monthly errors. Monthly precipitation tends to be a much smoother underlying function than daily precipitation. Over a month, precipitation almost always occurs at every station. However, for a single day most stations are likely to report no precipitation and only a few report large values when a significant storm passes through. This leads to the underlying dataset being very discontinuous. When monthly data are used for precipitation, errors tend to be on the order of 20% of the observed signal, much more reasonable than the daily values. Since we needed daily data for a lot of our statistics, we put up with the large errors. Also since the regional model errors were large, these errors in observations did not significantly impact any of our analysis.

II.4.3 Difference in Climatology

Figure 6 shows the temperature climatology produced for the Utah region. We see that the AE and LP techniques give almost identical answers. However, the AVG method gives a bias roughly 4 degrees Celsius from the other methods. This is significantly outside natural variation and is unacceptably large.

Figure 7 shows the precipitation climatology produced for the Utah region. For precipitation all of the methods are quite close to one another, but the AE and LP methods are obviously different. There is about a 5 mm difference between any of the techniques. In the dry months of June, July, and August the AE and LP methods are quite close, but the AVG method is slightly too dry by about 5 mm. For the Utah region all of the methods seem to work sufficiently well for precipitation.

II.5 DISCUSSION AND CONCLUSIONS

We have developed a new apple-to-apple comparison of the climate model to observations. We have met the statistical analysis challenge given by *Mearns et al.* [1997] to develop good statistical tools to evaluate climate model experiments. We have also established the limitations of some of the simple methods used to evaluate climate models. The methods we have developed could also be used in themselves to drive impact models of various sorts.

Our analysis compares the model to observations on the same grid and compensates for elevation differences. We have shown, as pointed out by *Giorgi et al.* [1993] that in the mountainous terrain of the US, the elevation difference of model grid-points to observed data-points is major and leads to a significant bias. We also adjusted the model temperatures to screen height. Typically this change increases the diurnal temperature range. The important thing to note is that it rectifies the meaning of the comparison of model to observe. This is because terms that have the same physical meaning are compared.

Our analysis used a suite of statistical analysis tools for understanding several different aspects of a climate simulation. What is the long-term model bias? How does the model bias compare to observed error? How does the model compare to using observed climatology as a predictor? Are the precipitation patterns realistic? And do these precipitation patterns compare well to the daily observational forecast? Does the model precipitation pattern forecast better than a random precipitation pattern? Lastly, what about the precipitation frequency distribution? If there is a model bias, is it because of an incorrect distribution of storm intensities, storm frequency, or storm duration? The suite of statistical tools we developed provides quantitative answers to these questions.

We found that a simple station average (the AVG method) over mountainous regions produced a significant bias in the results. The bias and the fact that the error estimates were significantly higher suggest that the AVG method not be used in evaluating climate simulations.

The LP technique could also be used for input into impact models. As we have defined it, the LP method could be used for current climate simulations, and the distributed data input into various impact models. Limitations of this method are limited resolution, restricted data-types, bounds on the simulation length, and difficulties associated with simulating climate change. Advantages include its low error statistics as compared to other gridding methods, the ease of use, and the high quality of error signal reported.

Because of the scarcity of data in mountainous western US regions, resolution of the LP method is limited if high accuracy is required. This method makes reasonable estimates for any resolution of data. At finer and finer resolutions the method relies more heavily on the regression model itself, rather than local data. By looking at the distribution of station data in Figures 2 and 3, we see that at resolutions finer than 50 km a majority of the grid-cells would not have data within or in nearby grid-cells. At this point changes in terrain (such as rain-shadow affects) or land-use changes can sway local values substantially away from the regressions. Consequently, the error estimation methodology will have to be changed slightly due to the lack of local data [*Lall and Bosworth,* 1994]. This will increase the error estimates for higher resolutions.

Likewise, the LP technique can be applied to other useful data-types (radiation, winds, humidity, etc.). The lack of observational data for these data-types (approximately a dozen observed stations

within our comparison regions) restricts its use to precipitation and temperature. Although the technique can be applied at a higher temporal resolution, the lack of high frequency data restricts such use. In some areas, the density of hourly observing stations (or stations that collect different data-types) may be dense enough to use LP, but definitely not in the barren regions of the mountainous US.

LP is also limited to estimating sequences when observed data are available. In the western US there is data back to the late 1800s, but it is very sparse. Reasonable 30 to 50 year data-sequences can be assembled, but not beyond that point. An advantage to the LP method is that the error estimates it generates will automatically show when the error is acceptable or not. Longer sequences will require the use of resampling schemes [*Lall and Sharma,* 1996] to come up with longer series of synthetic data. But, then the errors associated with the resampling come in to play, as well as the errors associated with the grid estimation. The main problem with resampling is that short time-sequences may not have the representative interannual variability that would be shown in a longer sequence. When resampling is used, a synthetic sequence is generated. This sequence cannot be directly compared with an observed sequence. Only climate statistics can be compared to observed statistics. Thus validation is much more difficult to do since only the long-term climate can be analyzed.

Adjusting LP generated climate data for climate change scenarios is a difficult task. Once again this requires some sort of resampling of observed data-sequences, and/or adjustment of climate statistics. This creates a simulation dependent on the worst features of both climate models and observed data simulation. The problem is that the method requires both the generation of synthetic data and requires analysis of a GCM simulation. Thus the biases inherent in both the synthetic data creation and the GCM simulation come in to play. On the other hand, the biases in physically based downscaling (see chapter III) can also be large. Thus, it is difficult to say what the best method for prognosticating future scenarios would be.

One problem with regional climate models is that they require a large expenditure of computer and personnel resources. And even for existing climate the errors can be substantial as we will see with RegCM2 in chapter III. Where good statistics for the data-types needed are available, the LP method does provide reasonable inputs with a much lower expenditure of resources than the regional model. The LP method has higher error levels for daily data, but reasonable error levels for monthly data. Hence, for many applications the use of a gridding technique such as LP or AE may be better suited for input into an impact model than using a regional model.

In summary, LP is a useful technique both for comparison of regional climate models and for creating distributed observed data-sequences possibly for input into impact models. LP shows smaller error estimates than other methods and values similar to the AE method. The AVG method has significant bias for these mountainous regions, and large error estimates and because of that we recommend against its use.

References

Barros A. P., and D. P. Lettenmaier, Dynamic modeling of the spatial distribution of precipitation in remote mountainous areas, *Mon. Weather Rev., 121,* 1195-1214, 1993.

Briggs, B. C., Machine contouring using minimum curvature, *Geophys., 39, 1,* 39-48, 1974.

Cleveland, W. S., Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.*, *74*, 829-836, 1979.

Cleveland, W. S., S. J. Devlin, Locally weighted regression: an approach to regression analysis by local fitting, *J. Am. Stat. Assoc.*, *83*, 596-610, 1988.

Cleveland, W. S., S. J. Devlin, and E. Grosse, Regression by local fitting, *J. Econometrics*, *37*, 87-114, 1988.

Cotton W. R., and R. A. Anthes, *Storm and Cloud Dynamics*, 883 pp., Academic Press, San Diego, Calif., 1989.

Deardorff, J. W., Dependence of air-sea transfer coefficients on bulk stability, *J. Geophys. Res.*, *73*, 2549-2557, 1968.

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, Biosphere-atmosphere transfer scheme (BATS) version 1E as coupled to the NCAR community climate model, *NCAR/TN-387+STR*, 69 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1993.

Dyer, A. J., A review of flux-profile relations. *Boundary-Layer Meteorol.*, *1*, 363-372, 1974.

Giorgi, F., Simulation of regional climate using a limited area model nested in a general circulation model, *J. Clim.*, *3*, 941-963, 1990.

Giorgi, F., and G. T. Bates, The climatological skill of a regional model over complex terrain, *Mon. Weather Rev.*, *117*, 2325-2347, 1989.

Giorgi F., and L. O. Mearns, Approaches to the simulation of regional climate change: A review, *Rev. Geophys.*, *29*, 191-216, 1991.

Giorgi, F., G. T. Bates, and S. J. Nieman, The multiyear surface climatology of a regional atmospheric model over the western United States, *J. Clim.*, *6*, 76-95, 1993.

Giorgi, F., C. S. Brodeur, and G. T. Bates, Regional climate change scenarios over the united states produced with a nested regional climate model, *J. Clim.*, *7*, 375-399, 1994.

Jacob, D. J., M. J. Prather, P. J. Rasch, R. -L. Shia, Y. J. Balkanski, S. R. Beagley, D. J. Bergmann, W. T. Blackshear, M. Brown, M. Chiba, M. P. Chipperfield, J. D. Grandpre, J. E. Dignon, J. Feichter, C. Genthon, W. L. Grose, P. S. Kasibhatla, I. Kohler, M. A. Kritz, Evaluation and intercomparison of global atmospheric transport models using ^{222}Rn and other short-lived tracers. *J. Geophys. Res.*, *102, D5*, 5953-5970, 1997.

Lall, U., and K. Bosworth, Nonparametric statistical inference and function estimation for hydrologic space time data, *Trends Hydrol.*, CSRI, 1-19, 1994.

Lall, U., B. Rajagopalan, and D. G. Tarboton, A nonparametric wet/dry spell model for daily precipitation, *Water Resour. Res.*, 32, 9, 2803-2823, 1996.

Lall, U., and A. Sharma, A nearest neighbor bootstrap for resampling hydologic time series, *Water Resour. Res., 32, 3*, 679-693, 1996.

Leung, L. R., and S. J. Ghan, A subgrid parameterization of orographic precipitation, *Theor. Appl. Clim.*, *52*, 95-118, 1995.

Luthi, D., A. Cress, H. C. Davies, C. Frei, and C. Schar, Interannual variability and regional climate simulations, *Theor. Appl. Clim.*, *53*, 185-209, 1996.

Macauley, F. R., *The Smoothing of Time Series*, 172 pp., National Bureau of Economic Research, New York, 1931.

Mearns, L. O., C. Rosenzweig, and R. Goldberg, Mean and variance change in climate scenarios: methods, agricultural applications, and measures of uncertainty, *Clim. Change*, *35, 4,* 367-396, 1997.

Press, W., *Numerical Recipes: The Art of Scientific Computing*. 702 pp., Cambridge University Press, New York, 1989.

Rajagopalan, B., U. Lall, and D. G. Tarboton, A nonhomogeneous markov model for daily precipitation simulation, ASCE, *J. Hydrol. Eng.*, *1, 1, 33-40,* 1996.

Rajagopalan, B., U. Lall, and D. G. Tarboton, Evaluation of kernel density estimation methods for daily precipitation resampling, *Stochastic Hydrol. Hydraul.*, *11*, 523-547, 1997.

Scott, D.W., *Multivariate Density Estimation*, 317 pp., John Wiley and Sons, New York, 1992.

Sheather, S. J., and M. C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, *J. R. Stat. Soc.*, *53, 3*, 683-690, 1991.

Silverman, B.W., *Density Estimation for Statistics and Data Analysis*, 175 pp., Chapman and Hall, New York, 1986.

Sumner, G., *Precipitation: Process and Analysis*, 455 pp., John Wiley and Sons, Chichester, Eng., 1990.

Wand, M. P., and M. C. Jones, *Kernel Smoothing*, 212 pp., Chapman and Hall, New York, 1995.

Table 1. List of Regions and Abbreviations Used for Each Region

	Region	Description
1	WA	Washington state area
2	WA-ID	Parts of Washington and Idaho
3	ID-MT	Parts of Idaho and Montana
4	OR	Oregon state area
5	ID	Central Idaho
6	CA-NV	Parts of California and Nevada
7	NV	Nevada state area
8	UT	Northern Utah
9	CO	Central Colorado
10	CO-NM	Parts of southern Colorado and New Mexico
11	AZ	Central Arizona

Table 2. Difference Between the Average Elevation for the Model Grid-Points Within Each Region and the Average of the Observed Station Elevations for Eleven (250 km)^2 Mountainous Regions. The Second Column is the Average of All the Model Grid-Points Within the Region. The Third Column is the Average of the Elevations for the Scattered Observation Stations Within the Region. And Finally, the Last Column is the Difference From Model to Observed

Region	Avg. obs. Elev.	Avg. model elev.	Difference
WA	1892	2204	312
WA-ID	2828	2415	413
ID-MT	3754	4014	260
OR	4359	4866	507
ID	4866	6014	1148
CA-NV	4623	5156	533
NV	5821	6268	447
UT	6364	6877	513
CO	8008	8815	807
CO-NM	8295	8581	266
AZ	4299	4508	209
Average	4592	4977	382

Table 3. Suite of Comparision Statistics

Statistic	Description	Purpose
Climatology	Average over each point in the region, and then average all years for each month separately.	Long-term model bias
Anomaly Correlation	The correlation of the time-series, of the monthly value minus the climatology, for the observed vs. the model	Time series correlation without the long-term signal
Model Skill	Score of the model bias vs. the bias when using climatology	Quantitative measure of how good or bad the model bias compared to using the Climatology as the predictor.
Precipitation Frequency	Frequency distribution of the daily precipitation for both model and observed	Estimate if the model has the right number of days between storms and storm lengths are reasonable
Precipitation Intensity Distribution	Frequency distribution of different daily intensities for both model and observed	Estimate if the model has the right distribution of both high-intensity and low intensity storms
Error	Comparison of the observed error to the model bias	How the model bias compares to the error generated using station data
Threat Score	Score of how good the pattern for a given threshold of precipitation matches the observed	How good the model precipitation forecast pattern is
Monte-Carlo Threat Score	Threat score minus the threat score for a random forecast pattern	Test, if the threat score is significant or achieved via random chance
Bias Score	Ratio of the number of model points greater than a threshold vs. the number of observed points greater than that threshold	Test if model rainfall areas are too small or too large

Table 4. Impact of the Data Screening on Precipitation Estimates over the Entire 14-year Period

Region	# of stations	Average # of stations used	Average difference (mm)	Max difference (mm)
1	232	85	-4.26	276.12
2	240	110	-2.84	45.7
3	219	103	-3.97	167.1
4	135	68	-5.09	158.85
5	194	103	-1.11	158.85
6	190	82	-18.24	2141.89
7	85	43	-3.17	193.44
8	238	121	-2.54	33.68
9	232	117	-19.78	2701.32
10	173	88	-5.58	2638.54
11	193	83	-24.24	2638.54
Average	178	84	-8	930

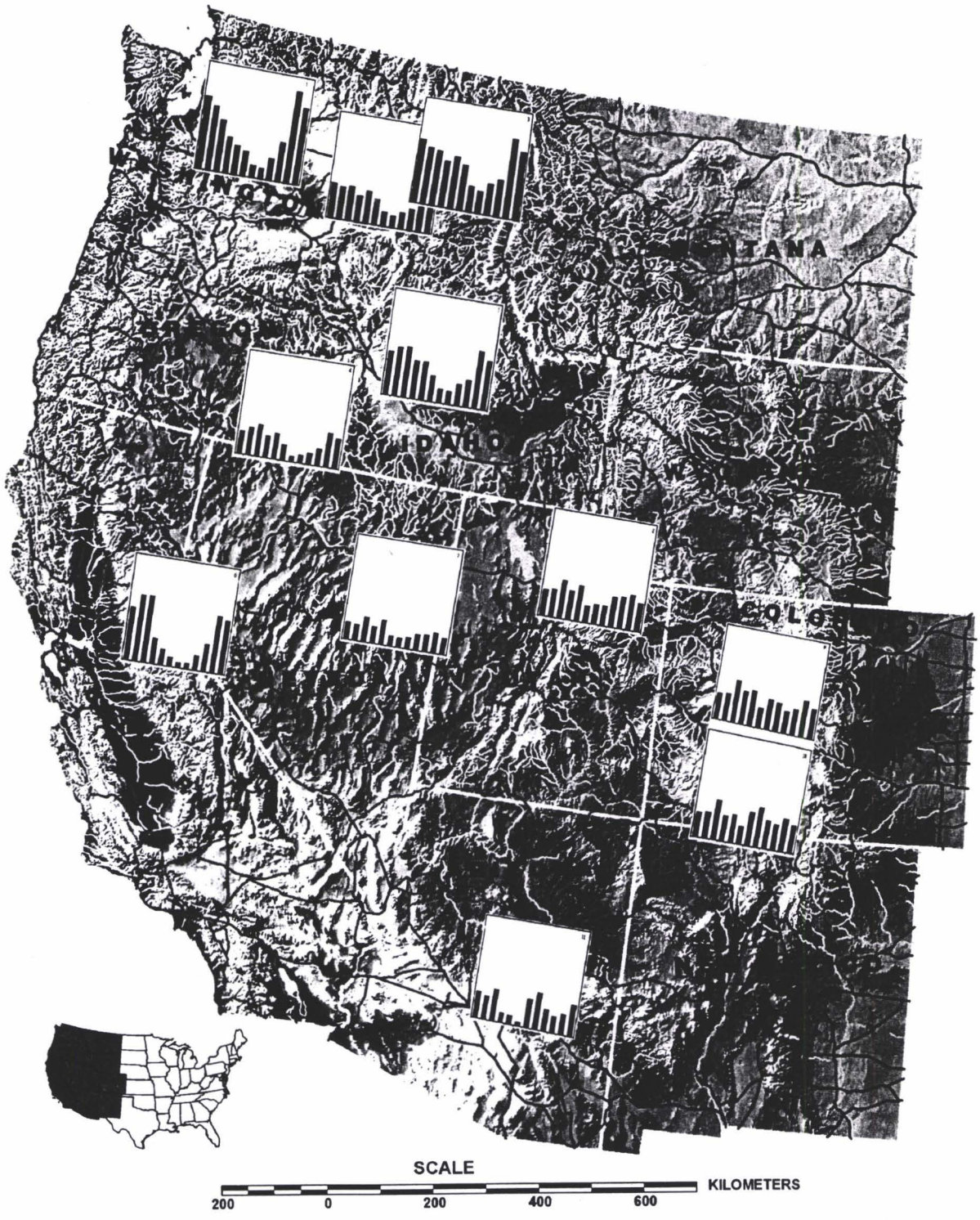

Figure 1. Eleven regions we use for comparison.

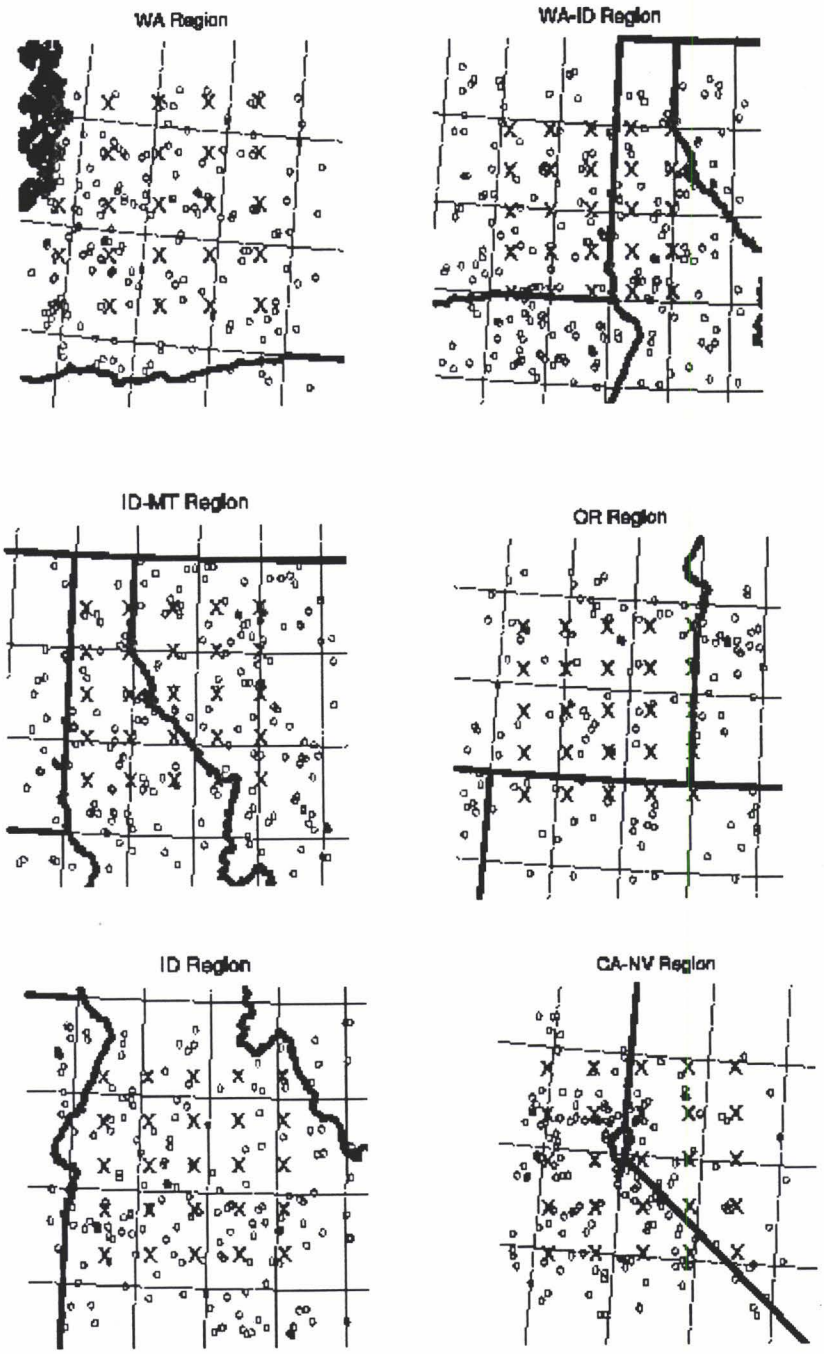

Figure 2. Location of the observed stations and grid points for six of the eleven analysis areas. Green circles are the station locations and red x's are the grid points.

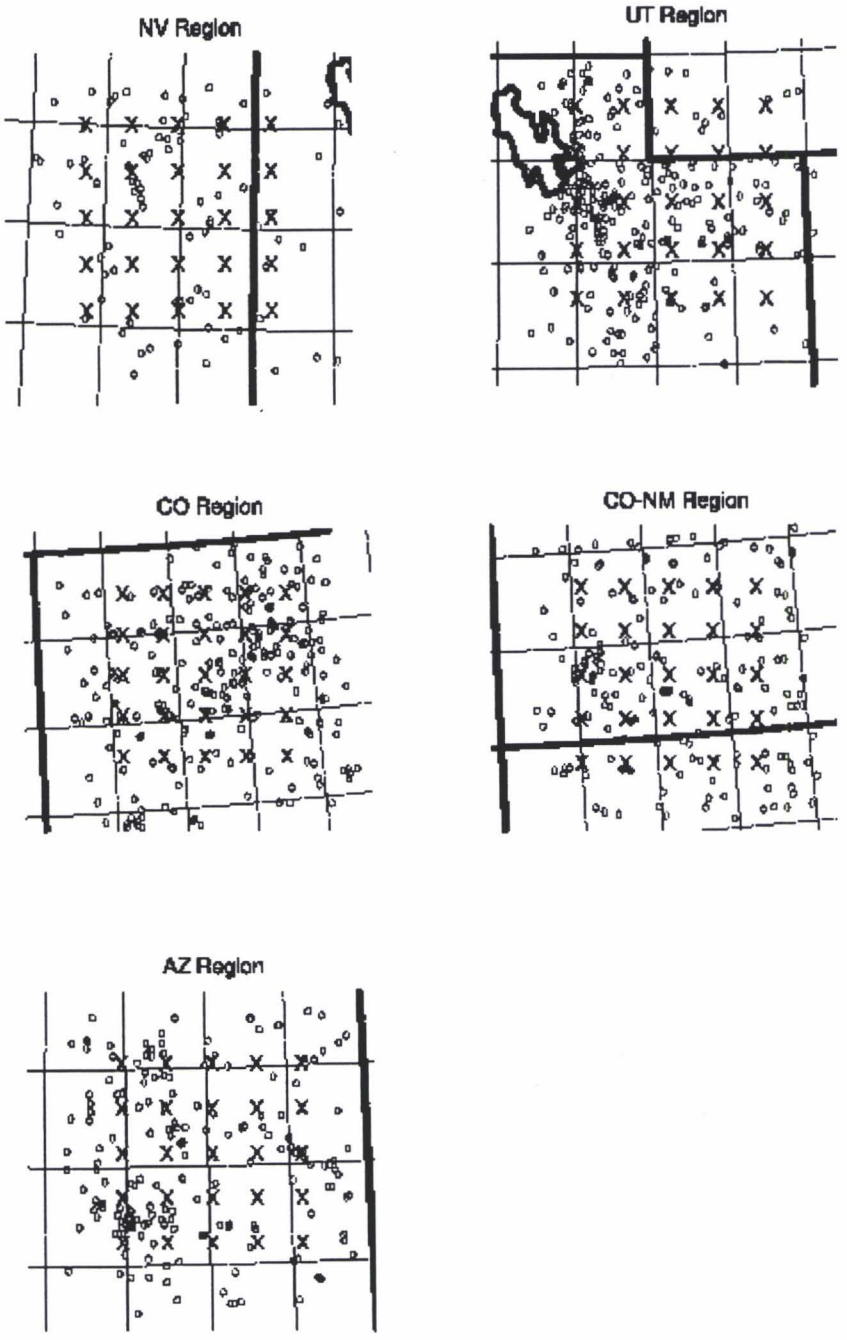

Figure 3. Location of the observed stations and grid points for five of the eleven analysis areas. Green circles are the station locations and red x's are the grid points.

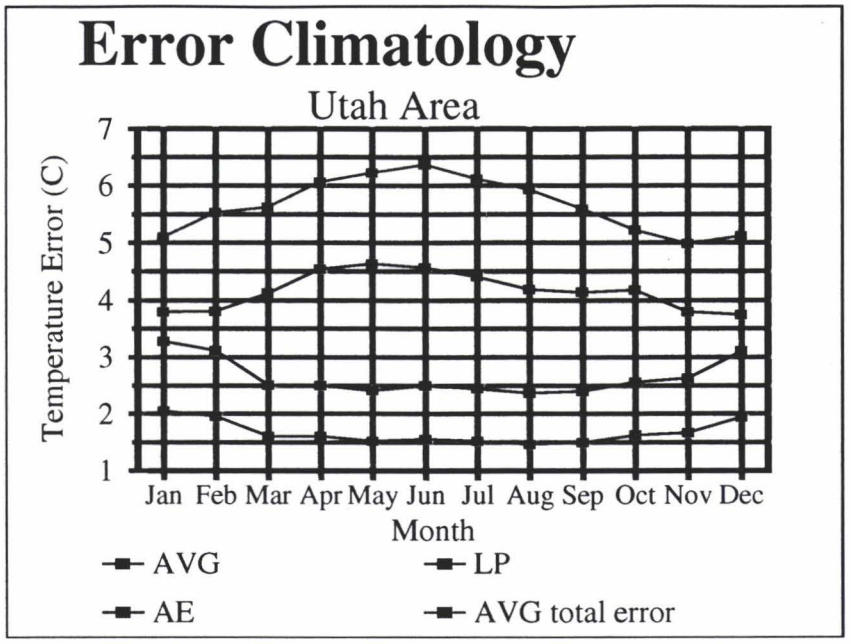

Figure 4. Comparison of T_{ave} error estimates.

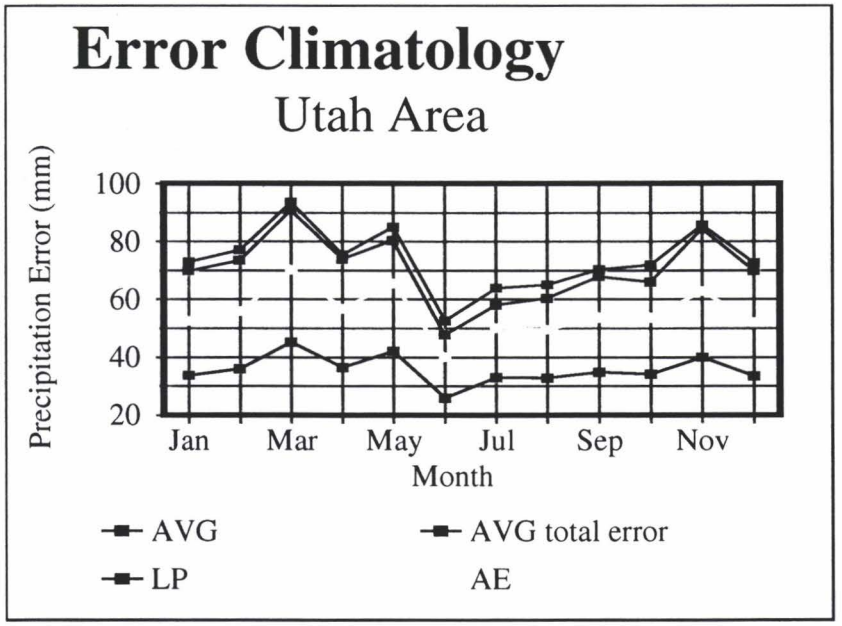

Figure 5. Comparison of error estimates for precipitation.

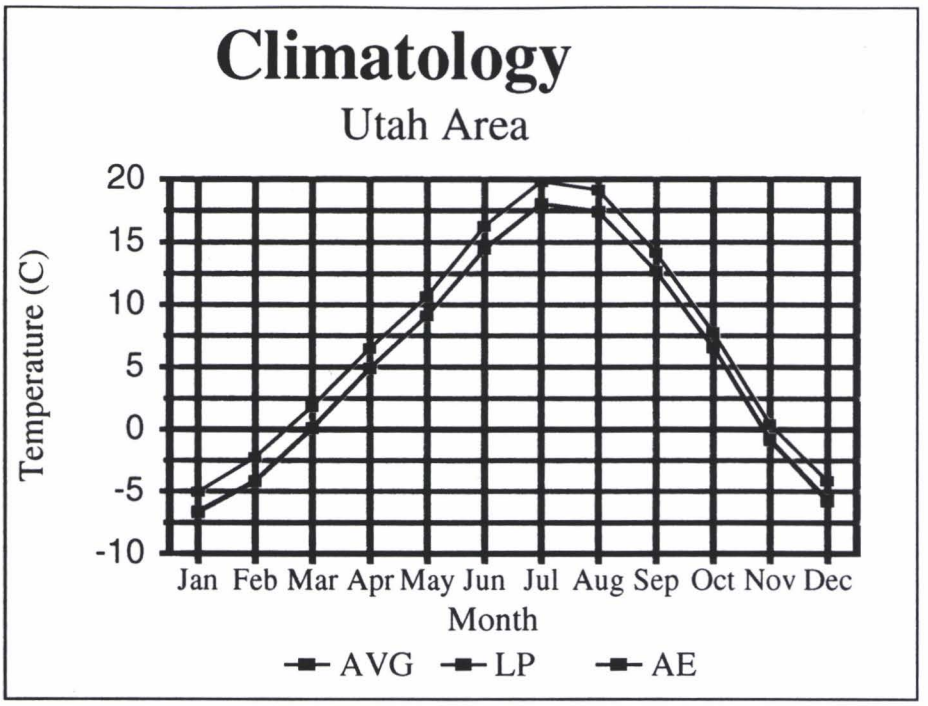

Figure 6. Period averages of average temperature for the AE, LP, and AVG analysis methods.

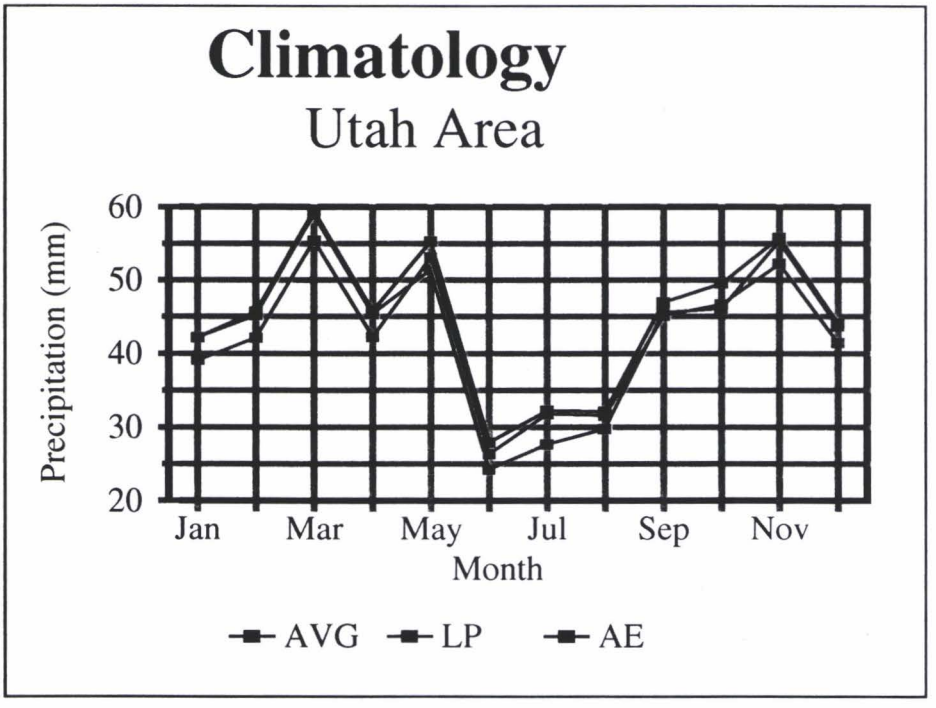

Figure 7. Period averages of monthly precipitation for the AE, LP, and AVG analysis methods.

CHAPTER III: DOWN SCALING GCM SCALE OUTPUT TO MOUNTAINOUS REGIONS USING A REGIONAL MODEL: PART II: LONG-TERM STATISTICAL COMPARISON

Abstract

Many groups use regional climate model data to drive ecological, hydrological, and surface interaction models. However, direct long-term comparisons of regional model skill are limited, especially in mountainous terrain. In this chapter, we use the methodology developed in chapter II to compare RegCM2 model output to observations at 50 km resolution. We use recorded climate station measurements in the western US over a 14-year span (1978 – 1992). This period includes multiyear cycles of wet and dry sequences, providing a good test of model capability. We compare the model to observed temperature and precipitation over monthly and period averaged intervals and examine daily precipitation frequencies and patterns. This provides a well-rounded analysis for investigating the capability of the regional model. We apply these statistics over eleven 250 km^2 mountainous regions distributed across the western US.

III.1 INTRODUCTION

The study of the potential impact of enhanced greenhouse warming on natural systems and economic activities (such as agriculture) is experiencing some lag in support. Conversely, global and regional studies of climate–land surface responses are expanding rapidly. Over the past decade, careful intercomparisons between different global climate models (GCMs) [*Chen et al.*, 1997; *Fulker*, 1997; *Gaffen*, 1997; *Gaffen et al.*, 1997; *Min*, 1997], GCM simulations, and current climate and perturbed climate (climate change) scenarios have led to vast improvements in model accuracy. The need to use GCM outputs for regional and local scale problems has fostered a broad effort in scale adjustment. Additionally, climate inputs are often required at daily time scales to evaluate the potential impact of human activities such as agriculture, forest utilization, and water storage. This problem is accentuated in the mountainous regions of the western US because of the wide differences in terrain and climate and the scarcity of observational data.

When applying downscaled GCM data to vegetation and hydrologic studies, the short-term variations in temperature and precipitation are often lost in climatic averaging [*Paoli*, 1994]. It is often these fluctuations that provide the stresses that limit vegetative development and cause extreme events. Observed station extrapolation to surrounding regions works well for many studies, often retaining the observed extremes of a particular location [*Woolhiser and Roldan*, 1982; *Wilson and Lettenmaier*, 1992; *Bardossy*, 1994; *Turner et al.*, 1996; *Dodson and Marks*, 1997; *Rajagopalan et al.*, 1997]. Applying these methods in complex terrain requires special care. A limitation of these methods is that they only simulate observed variation. Resampling schemes can be used to shift the frequencies of extreme events but an understanding of GCM implied shifts in variability must be developed by some other method [*Paoli*, 1994] (see chapter II).

Several methods have been developed to apply GCM scale data to local studies [*Zhang and Foufola-Georiou*, 1997]. Changes or extremes in local climate may not be detected by methods that extrapolate from existing observed and averaged GCM outputs or that apply empirical transfer functions [*Palutikof et al.*, 1997]. Nested climate modeling has been suggested as a physically

based approach to correct for the sub-grid effects ignored by the larger scale models [*Avissar and Pielke,* 1989]. However, full downscaling using this approach comes with a major cost in computer resources and time. Significant differences have been found between GCM and regional scale models, and observed data [*Portman et al.,* 1992; *Privalsky and Croley*, 1992; *Copeland et al.*, 1994]. Reviews of the significance of land surface schemes in the models show that these input variables require careful attention. *Leung and Ghan* [1995] show that a high-resolution sub-grid-scale representation can improve surface variables better than a lower resolution (but more computer intensive) nesting down approach. Thus, the sub-grid surface fluxes and orographic effects may significantly impact the model simulation in arid mountain regions [*Leung and Ghan*, 1995]. This shows the need for higher resolution modeling simulations.

Accurately comparing regional model outputs in mountain areas with observed data requires major effort [*Giorgi et al.,* 1992, 1994; *Barros and Lettenmaier*, 1993; *Leung and Ghan,* 1995]. The few existing studies of regional model downscaling in mountain regions rely on short-term (1-2 year) comparisons. Clearly, a major investment is made in yearlong simulations at this scale. However, annual variability among wet and dry years can still result in wide variations in these comparisons [*Luthi et al.*, 1996; *Adams and Comrie*, 1997; *Battisti*, 1997]. This is particularly true over areas the size of the western US. *Giorgi et al.* [1993a] published a comparison between RegCM output and 390 National Oceanic and Atmospheric Administration Cooperative observation network (NOAA COOP) stations distributed across the western US. Our study compares a 14-year model simulation to observed data sequences of the same length and period. This allows annual variability to be examined.

The analysis presented in this chapter investigates the capability of a regional model to provide a physically based system to link global decade-scale climate inputs to vegetation change and river basin flow studies. We demonstrate a rigorous statistical comparison of the model output to observations, using the methodology developed in chapter II. In this study we are particularly attentive to the quality of the model and observed grid outputs to be used as input data for regional, watershed, and local scale models. Using our suite of comparison statistics, we will demonstrate some particular weaknesses in the RegCM2 model simulation. As we will see, the model has significantly large biases. If the long-term biases are compensated for, the model demonstrates simulation skill. Therefore, if the model can be tuned to eliminate the long-term biases, the simulations will be useful. However, the model biases are currently too large for utility in most impact models and are significantly outside natural variability. The model shows utility for climate change scenarios if biases for current climate are assessed at the same time, and those biases are accounted for in the conclusions of the study. In order to assess model performance for different areas we select several mountainous regions to compare to observations. We also detail the specifics of the model and the simulation characteristics. In the results section we will compare model to observations for each analysis region for the statistics outlined in chapter II, which assess the performance of surface screen-temperatures and precipitation. We find significant variation in the model performance depending on the region; some regions perform significantly better or poorly than others. Also there is a significant variability in model performance month to month and even year to year. In light of this variability and in order to properly evaluate the regional model, more decade-long simulations need to be performed. Many of the statistics we show would be inconclusive for a shorter simulation period. A spin-up time of at least a year is needed for the temperatures, and underscores the need for longer simulation periods. To date our simulation is the longest such simulation with a regional model of which we know.

III.2 STUDY AREAS

The western US provides an interesting study area for GCM downscaling. It contains several mountain ranges that are not resolved at the GCM scale, such as the coastal ranges, the Sierra Nevada, the Cascades, the Wasatch, the Bitter-Root, and the Front Range of the Rocky Mountains. This area is large enough to support significantly different regional climates. This arid region also depends heavily on wintertime mountain snow-packs to support its agriculture enterprises and urban populations. Figure 1 shows a Digital Elevation Model (DEM) of the western US, with our eleven 250 x 250 km square analysis areas outlined and the observed precipitation climatology reported for each. Data from the model are compared with data from the manual and automatic climate stations in these same regions.

We compare small, localized regions, but not to any particular point. In assessing how best to make this comparison, it was noted that few mountain watersheds extend beyond the 5 grid-cell (250 x 250 km) dimension. Mountain watersheds across the western US were chosen to allow analysis of the model's behavior across a range of climatic and synoptic scale differences. It was decided to use a box of 5 x 5 grid-points for each comparison. This way patterns could be examined, data could be averaged over several model points, and very fine scale regions could still be investigated.

Figure 1 shows that for most of these regions, the 250 x 250 km square block represents the core of the mountains contained within the area. In most cases, the area includes both sides of the mountain range(s) and therefore includes rain shadow effects. These effects are evident in the error analysis presented later in this report.

We examine our gridding methods over 11 small 250 x 250 km square regions spread throughout the western US. Most of the examples we show are for one of the regions, northern Utah. These regions are the same regions compared to the model in chapter II. In Table 1 we list the regions and the abbreviations we will use in addressing them.

III.2.1 Regional Atmospheric Model: RegCM2

The RegCM2 [*Giorgi et al.*, 1993b, 1993c] model is a hydrostatic, compressible, sigma layer, primitive equation atmospheric model. The basic model is mesoscale model version 4 (MM4) as described by *Anthes et al.* [1987]. Although a more recent mesoscale model for numerical weather predication has been released (MM5), NCAR developed RegCM2 from MM4 specifically for climate studies. The advancements to MM4 to make MM5 were: a non-hydrostatic option, ice microphysics, and a sophisticated nesting scheme. None of these options are desired for climate simulations since they are all inherently CPU-intensive. However, advancements in the radiation, surface, and boundary layer physics are all required for long-term climate studies. The NCAR climate-modeling scientists wanted to implement these advancements to reduce CPU time and errors in the pressure-gradient calculation. Therefore, NCAR took MM4 and added sophisticated packages for surface physics (Biosphere Atmosphere Transfer Scheme version 1e (BATS1E)) [*Dickinson et al.,* 1993], lakes [*Hostetler and Giorgi,* 1992], and a non-local atmospheric boundary layer [*Holtslag et al.*, 1990 and *Holtslag and Boville*, 1992]. A delta-eddington approach to solar radiation [*Briegleb*, 1992], and an option to use the Grell convective parameterization scheme [*Grell et al.*, 1994] were used with an *Arakawa and Shubert* [1974] closure. RegCM2 is

an improved version of the RegCM model used in several earlier studies [*Giorgi and Bates*, 1989; *Giorgi,* 1990, 1991; *Medina*, 1991; *Giorgi et al.*, 1992; *Matthews et al.*, 1992].

To save CPU time, RegCM2 uses a split-explicit time differencing approach [*Madala*, 1981] rather than the explicit time-integration approach of Brown-Campana described in *Anthes et al.* [1987]. This approach allows use of a time-step more than twice as large without loss of accuracy compared with the Brown-Campana technique. Also the pressure gradient formulations of *Corby et al.* [1972] are used to reduce numerical errors in the pressure gradient term along sloping sigma-surfaces. In this study, the model is applied to a similar area to that studied by *Giorgi et al.* [1992] except that a grid domain of 72 x 60, 50 km square boxes, with 17 vertical layers is used. We also apply terrain smoothing on the lateral boundaries, similar to *Hong and Juang* [1998]. The lowest model layer is approximately 20 m off the surface and the top model layer is 80 hPa. A contour plot of the model topography used in this simulation is shown in Figure 8. The difference between the smoothed 50 km model terrain and actual terrain can be easily seen by comparing this contour plot of the west with the digital elevation model (DEM)-based terrain depicted in Figure 1.

We used the RegCM2 regional model to create a simulation at 50 km resolution over a 14-year period from December 1978 to December 1992. This period included multiyear cycles of both wet and dry sequences, providing a good test of model capability. Boundary conditions for our simulation were derived from the *Trenberth and Olson* [1988] National Center for Atmospheric Research (NCAR) archive of the European Centre of Medium Range Weather Forecasting (ECMWF) observational analysis. NCAR stored the data at T42 (2.81 x 2.81 degrees) resolution even though the resolution of the original analysis varied during the sequence. This was done because NCAR wanted to compare the ECMWF analysis archive to the T42 simulations of the community climate model version 2 (CCM2) GCM. The datasets were acquired from the NCAR mass storage archives and were similar to the input data used by the Giorgi simulation [*Giorgi et al.* 1993a, 1994]. The ECMWF analysis is an operational product and the quality of the analysis changes throughout the series [*Trenberth and Olson,* 1988]. The initial part of the simulation used ECMWF data at 12-hour updates. But after 1986, 6-hour update data were available and were used in the simulation. Observed sea surface temperatures (SST) for the Pacific Ocean were developed by interpolating the global Climate Analysis Center (CAC) 2° monthly SST dataset for 1978 - 1981 [*Reynolds*, 1988]. After that time, we used the CAC blended dataset. This dataset blends in situ data, composed of ship and buoy measurements, and satellite data to form a gridded dataset [*Reynolds and Marsico*, 1993]. The soil moisture initial condition was set at the midpoint between saturation and wilting point, the same as done by *Giorgi and Bates* [1989]. Initial snow-cover was developed from the December 1, 1978 observational data, with snow-water equivalent set to 10% of the reported snow-depth, and surface temperature of snow-covered areas set to freezing. Even with these starting conditions, most of the snow-cover still melted almost immediately, not reappearing until January. It was not until the next winter-season that we had realistic snow-cover. Modeled output was quality controlled on a daily basis using the hydro-meteorological data video display system developed by *Oborn et al.* [1994].

Because gridding techniques are used to bring observed precipitation and temperatures to the model grid position and elevation, both datasets are directly comparable. See chapter II for more information on the gridding technique used to produce estimates at model grid-points.

III.3 RESULTS

We now present the results of our comparison with the observed using our suite of comparison statistics (Table 3). Climatology plots shows that while mean temperatures are modeled well, the diurnal temperature ranges are significantly too low by 6 to 10 degrees Celsius. Also, in general, precipitation is much too high especially in the summer. In many cases precipitation is double the observed, a value well over natural variability. However, the anomaly correlation and skill plots show that after these long-term biases are taken out, the model does show simulation skill. Some regions are modeled better than others are and there is a wide variation of skill depending on the year. For some regions and periods, even with the long-term bias extracted out, the model performance is unacceptably poor. The threat score and Monte-Carlo threat score both show that the model does a reasonable job of capturing the timing and patterns of observed storm events. For thresholds much above zero, however, the model quickly shows little or no performance. The bias score shows that the model precipitates over areas from about 0 to 50% larger than observed. The frequency distributions show that the model in general does a good job of simulating the distribution of precipitation intensities. However, for most locations the model tends to have precipitation too often and for too long.

III.3.1 Model and Observed Climatology

We first present the period averaged monthly results for temperature and precipitation for each of the analysis areas. The comparisons of the model and observed maximum monthly air temperatures are shown in Figures 9 and 10 for each of the 11 analysis areas. In the Washington and Oregon areas, which are at a lower elevation, only a small difference is shown. However, in most of the other regions, the model underestimates the monthly average of the daily maximum temperature by 2 to 3 degrees Celsius for most of the year. This is a significant bias well above the normal variability in the observed.

Comparison of observed and modeled monthly averaged daily minimum air temperature is shown in Figure 11 and 12 for each of the 11 analysis areas. In all locations, except Arizona, the model minimum monthly temperature is 2 to 3 degrees Celsius warmer than the observed data. In the Idaho area this difference increases in the summertime, with the average difference approaching a full 5 degrees Celsius for July and August. During the summer, minimum temperatures for Arizona are not significantly different, but the two curves separate for the winter months. Once again the model bias is significantly outside normal variability of the observed.

Because the model biases in the maximum and minimum temperatures are nearly equal and have opposite sign, the mean temperatures compare very well to observed. Comparison of observed and modeled mean monthly air temperature is shown in Figure 13 and 14 for each of the 11 analysis areas. In all locations, except Arizona, the model's mean monthly temperature is not significantly different from observed. The curves separate in the Washington, Idaho, and California areas for the spring months, but significant overlap remains in the observed error estimates. The model simulates the average temperatures appropriately, but the diurnal temperature range is much smaller than observed.

Model and observed monthly precipitation climatologies and standard deviations for the 11 analysis areas are shown in Figure 15 and 16. In almost all of the areas, the model overestimates precipitation by a factor of almost 2. In many of the regions the annual shape of the precipitation

curve is preserved, but offset and in cases slightly exaggerated. In the California area, the wide variation in the spring months makes the two curves not statistically different, but they separate during the summer months. Again the model overestimates the observed precipitation by 3 times. In the Colorado areas, where the highest mountains in the model terrain are located, summer precipitation is overestimated by 3 times also. However, in these cases, the shape of the curve differs significantly during the high sun months. In the Arizona region, the model picks up the summer monsoon in July, but does not reproduce the preceding dry May – June period. Again a bias of about 2 times is observed during the summer period.

Model and observed cumulative monthly climatologies for the 11 analysis areas are shown in Figure 17 and 18. These graphs show the annual differences in total precipitation and monthly precipitation amounts for the 14-year period. Despite the high model bias a wet/dry year for the model clearly corresponds to an observed wet/dry year. Likewise, although not shown here, a hot/cool year for the model corresponds to a hot/cool year for the observed. However, the model typically does poorly in the summer months with too much precipitation. In the California-Nevada area, the model actually underestimates the precipitation for some of the years.

Summary comparisons of the model output to the observed climatology, as shown in Figure 19, show the average errors associated with the climatology plots in Figure 9 through Figure 16. It is instructive to compare these error estimates with the annual average model bias for each element that is shown in Figure 20.

The annual average temperature error estimates (Figure 19) in each of the regions range from about 1 degree Celsius for maximum air temperature in Arizona to about 2 degrees Celsius for the minimum air temperature in five of the regions in the center of the model domain. It is interesting to note that the error estimate is almost always larger for the minimum temperature than for the maximum. The mean error estimate falls between these values, since it is computed from the two initial series. The annual average temperature bias between the model and observed (Figure 20) curves for all of the areas is about 3 degrees Celsius, and would have a substantial effect on plant growth and surface hydrology if not calibrated out of these models.

Anomaly is the difference between the period average climatology and the annual monthly values. Figure 21 shows the correlation coefficient between the model anomaly and the observed anomaly (for more information on the anomaly correlation see chapter II). The temperature anomaly correlation is very good, with correlation coefficients in the 0.8 – 0.9 range. A box-by-box comparison of the precipitation error and the model bias is instructive. The model bias toward heavy precipitation ranges from 18 mm over Utah, to 47 mm for the Washington-Idaho box. The significance of the models wet bias is appreciable. In almost all cases, the model's wet bias comes close to doubling the area annual average precipitation. The observed precipitation error based on the monthly data is at a reasonable uncertainty of about 20 percent the annual totals. However, as shown in chapter II, the errors for the daily data (Figure 19) are even worse than the model bias and around three times the annual totals. The monthly data remains suitable for comparison to the regional model. Impact models requiring daily data are served best if data is adjusted so that precipitation totals and temperatures averages agree with values computed with monthly data.

Despite the bias, the model versus observed anomaly correlation remains significant for all areas. This correlation ranges from 0.56 in the Colorado-New Mexico area to 0.8 for Idaho. At the 95 percent confidence level the correlation is significant if its value exceeds 0.15. This implies that

for all area the model anomalies are correlated to the observed anomalies. Hence, the model will properly show a wet year as wet, and a dry year as dry.

Even though the model has a significant bias, it is able to predict a wet year and a dry year. If the model is compared to observed, we see the large bias as characterized by Figures 15 and 16. However, if the model is compared to itself, it is able to determine which years are wet and which are dry. Although the model has a significant bias, the results still show predictive skill, provided the long-term bias is accounted for.

III.3.2 Temperature and Precipitation Forecast Skill

In this section, we examine the skill and bias of the model compared to the observed estimates of temperature and precipitation using daily input data, rather than the monthly results. The skill (for more information see chapter II) compares the bias removed model monthly anomaly with observed climatology. Because the skill varies greatly from month to month, we plot the skill by finding the running median of the preceding 12 months. When the model and the observed data are the same, the skill score will be 1. However, these charts are not quite as easily interpreted because two factors can affect the score. The skill is dependent on both the size of the median monthly anomaly, and the adjusted model bias. At a skill of zero, the bias adjusted difference is equal to the median monthly anomaly. Hence, the model does no better than using the observed climatology as the predictor. Values above 0 show the model observed difference is less than the median observed anomaly, so the model performs better than the climatology. Negative values are months where the model bias is larger than the monthly anomaly, so the model performs worse than the climatology. A little complex at first glance, these plots carry a lot of information about the model behavior.

The temperatures were relatively well behaved so we show only one set of area plots – the model maximum temperature predictive skill scores (Figures 22 and 23). Initialization error can be seen in the first year of the model, but then for most areas the model shows some simulation skill with skill scores around 0.5. A skill of 0.5 implies that the long-term corrected model bias is roughly half of the average error using climatology as predictor. There are periods for some areas, however, that show breakdown in the model's capability. For example, in the Colorado-New Mexico area during the 1985 summer, the skill goes negative. This indicates that a simple climatology approach would have provided better results for this period. A similar period occurs in the Arizona analysis for late 1992. In general, while skill is shown across all of the areas, some years appear to be more difficult for the model than others. Annual average skill levels are shown in Figure 24. The averages were computed without the first year, to eliminate the model spin-up errors in temperature. The model average skill levels, as displayed here, show a higher skill level in the northwest, with the lowest occurring in the Arizona analysis area. Because the observed minimum temperatures in mountain regions often involve inversions, the minimum temperatures often show higher errors. This factor may contribute to the lower skill level shown by the model in predicting the daily minimum temperatures.

Figures 25 and 26 show the precipitation skill score. These plots were generated using monthly data (for more information on the skill score see chapter II). Figures 25 and 26 clearly indicate that some regions are better predicted than others are. Washington, Idaho-Montana, California-Nevada, Idaho, Utah, and Arizona are all modeled with skill. Washington-Idaho and Nevada are not.

Oregon, Colorado-New Mexico, and Colorado are modeled at about the same level as the climatology over the 14-year period, with some years better and some worse.

The model precipitation performance requires a more complete analysis. To help us understand the model's lower performance in providing good precipitation information in these mountain regions, we first computed the model precipitation bias score (BS) (for more information on the bias score, see chapter II). The BS is the ratio formed from the number of the 25 model points for each month, which the model predicts above a threshold divided by the number of observed points above the threshold. If the two agree, the bias is 1. If there are more model points than observed above the threshold, the bias score is greater than 1. Conversely, if there are more observed points above the threshold, the bias score is less than 1. Figures 27 and 28 show the bias scores for each of the areas. Notice that for small thresholds, 0 to 1.5 mm, the bias scores range from 1 to 1.5 in most areas. This indicates that the model typically rains over an area 0 to 50 percent larger than observed. In the Washington-Idaho, Idaho-Montana, Idaho, Oregon, and Colorado-New Mexico areas, the BS is around 1.5 or greater and hence precipitation occurs over an area about 50 percent larger than observed precipitation patterns. In Nevada, the BS is near one, hence the model matches the observed precipitation areas quite well. In all cases, the model is under predicting the larger thresholds. In the Washington area, the BS is about 0.7 for 5 mm storms and at 0.5 for 10 mm storms. However, in some regions, such as Utah and Arizona, the bias score for 10 mm storms drops as low as 0.1. The bias score is undefined when there are no observed points greater than the threshold, but zero when observed points exceed the threshold while no model points do. Thus a small BS does not necessarily indicate that the model rarely has storms exceeding the threshold, but that they seldom occur at the same time that the observed exceeds the threshold.

The precipitation threat score (TS) (see chapter II) for each of the areas is shown in Figures 29 and 30. The TS is often used in forecasting to evaluate how good a model is for a given precipitation threshold and location. TS is also used effectively by *Giorgi et al.* [1993a] for use with regional climate models to show the predictive skill of these models. The threat and bias scores help to verify that not only are the total amounts of precipitation correct in the region, but also that the precipitation pattern within the region is correct. They also ensure that the simulation is reasonably similar to the observed time series. This is calculated for each day of the simulation. Here the data are calculated on a daily basis, and plotted with a 30-day moving average. The TS and BS are similar in that they are both calculated on the basis of how many of the grid-points are above the threshold. It is important to note the strong increase in model accuracy in the winter and early spring, with a significant decrease in skill in the convective seasons of late summer and fall. Again, we see the lowest skill levels with the higher precipitation events. The dark line in the figure is the 0 threshold with a 1-year moving average, which provides a good indication of how the TS varies on an annual basis. Some of the areas show some structure (like Nevada in the years 1983-1985) but most are quite flat. Again, the model appears to have the most accuracy in the northern areas, and the least in Arizona.

As a comparison for the model skill levels we examined what kind of skill levels would be exhibited with a random precipitation generation scheme. Is the model TS significant enough that it would outperform a random scheme? We developed a Monte-Carlo threat score (TS_{mc}) in chapter II. The TS_{mc} generated data for each point by randomly sampling some other day of the month for each point and then taking the average TS produced by a large sampling of this random data. The TS_{mc} is shown in Figures 31 and 32. The TS_{mc} is defined as the model TS – TS

(random) so that if the TS_{mc} is greater than 0, the model is outperforming the first-order stochastic model. In all areas, the model significantly out performs the random generator for low thresholds. The higher thresholds, however, show a much smaller chance at being significant on that day. The numbers in the boxes below the lines show where the curve drops below zero (stochastic procedure has more skill). The California-Nevada, Colorado, and Colorado-New Mexico areas show parts of several years where the model does not handle the precipitation better than a stochastic generator would. These periods are all in the high sun areas, where most of the precipitation is being generated by the convective parameterization in the model.

Because of the large annual cycle in some of the bias and threat score graphs, it is instructive to show these relationships on a climatologically averaged basis. Figures 33, 34, and 35 show the annual summary of the precipitation statistics. Figure 33 shows the annual average bias score. Since this is the ratio of the model over the observed data averaged over each point in the analysis box, a value of 1 shows equal response. Values for 0 through 1.5 mm are almost identical to each other, while values of 5 mm and above show a significant decrease. The threat score in Figure 34 shows a similar pattern, with good predictive capability for the lower intensity events and a lower score for the larger events. Figure 35 shows the TS_{mc} for the average performance of the model over the simulation period. The figure quantifies the model scores showing good average threat scores in the regions where substantial portions of the precipitation is associated with organized weather systems. But it shows negative scores in Arizona and Colorado where the precipitation is dominantly convective.

III.3.3 Precipitation Frequencies

In the analysis to this point, we have shown that the modeled precipitation is strongly biased toward levels at about twice the amount that is produced by a rigorous observed data gridding technique. This is within the error levels of the observed technique, but it is instructive to figure out how the modeled precipitation series differs from the observed. In addition, the model shows even larger summer time precipitation in some areas, but it seems to predict well for small events and with lower skill for events larger than 5 mm. In older versions of RegCM2, much of the problem with excessively wet summer time periods was attributed to "gridpoint" storms associated with numerical instabilities [*Giorgi*, 1991]. In this section we examine the cause of the differences between the model and observed data time series using the frequency and intensity analysis developed in chapter II. The model versus observed data difference will arise from a mismatch in the number of storms, in the length of the storms, or in the intensity of the storms. A mismatch in number of storms will show up in the dry spell frequency. A mismatch in the length of the storms will show up in the wet spell frequency. A mismatch in the intensity of the storms will show up in the intensity frequency. In the following section, we examine the difference between the modeled and observed storm intensities. In the section after that, we will examine the storm frequencies. In calculating these frequencies, we arbitrarily set a threshold for a rainy day. Since grid-scale averages (both model and observed) can have extremely small rainfall levels, we chose a threshold of 0.025 mm/day as the wet-dry day cutoff. This is a tenth of the smallest measurement of a modern observation station gauge. The observed series is the average precipitation over the entire 250 x 250 km area so if 100 gauges exist within the area and only one reports 0.25 mm, the average precipitation level will be roughly 0.0025 mm. Thus the grid average value can be much smaller than the smallest measurement of a single station.

III.3.4 Precipitation Intensity Distribution

The winter (December, January, and February) precipitation intensity distributions are shown in Figures 36 and 37. These probability density functions show that for these large areas, the events are strongly biased toward the minimum measurable event. The cumulative distribution curves show that in four areas, Washington, Washington-Idaho, Oregon, and Nevada the model intensities are slightly heavier than the observed stations. In four other areas (Idaho-Montana, Idaho, California-Nevada, and Arizona) the reverse is true. In the remaining three areas (Colorado-New Mexico, Colorado, Utah), the model and observed distributions are nearly identical. In a majority of the areas, the peak in the observed intensity curve is slightly higher and a little to the left of the model peak. In all cases, the differences in the distributions are quite reasonable, and we probably cannot attribute the large model precipitation biases that we have observed to differences in intensity.

The summer (June, July, and August) precipitation intensity distributions for the analysis areas are shown in Figures 38 and 39. These probability density functions again show the anticipated peak distribution strongly biased toward the minimum measurable event. This is not unreasonable for the large areas being compared. The cumulative distribution curves show that in eight of the areas (Washington, Washington-Idaho, Idaho-Montana, Oregon, Idaho, California-Nevada, Nevada, and Colorado) the model intensities are slightly lower than the observed stations. Indeed, in the Colorado and Idaho-Montana areas, the bias is rather large. The reverse is true in only two of the areas (Utah and Arizona). Only in the Colorado-New Mexico area are the model and observed distributions nearly identical. In a majority of the areas, the peak in the observed intensity curve is slightly higher and a little to the left of the model peak. In all cases, except Colorado, the differences in the distributions are quite reasonable, and we most likely cannot attribute the large model precipitation biases that we have observed to differences in intensity. In the Colorado area, the model clearly underestimates the intensity distribution.

III.3.5 Wet and Dry Spell Analysis

Since the intensity distributions do not show significant differences between the model and observed data, we are left to conclude that a significant difference remains to be observed in the number of wet days per storm and the number of dry days between storms. The winter (December, January, and February) wet spell analysis of our two primary data series is shown in Figures 40 and 41. The close match between the cumulative precipitation curves shows that for most areas the model does a good job of simulating the wintertime wet spell periods. The model seriously under estimates the wet spell periods in Oregon, and somewhat under estimates those in California-Nevada and Utah. On the other hand, it over estimates those in the Washington-Idaho region. Because the model slightly under estimates intensity in Oregon, the two effects cancel. The underestimation of intensity, coupled with overestimation of storm lengths, partially compensates for each other. So the resulting total bias is not quite as large as it could be.

The summer period (June, July, and August) wet spell analysis is shown in Figures 42 and 43. Here, we finally begin to see the cause of the strong model bias that results in nearly twice too much precipitation in all regions, and a heavy precipitation bias in the summer. In those regions where no strong summer bias was observed (Washington, Washington-Idaho, Idaho-Montana, Oregon, and Idaho), the wet spell lengths and distributions fall right on the observed distributions.

The rest of the regions have strong radiation forcing and some monsoon flow, and the wet spell periods are overestimated. In Colorado and Arizona, only 90 percent of the cumulative distribution function (CDF) weight has been accumulated after 30 days. Part of the problem in this comparison may be the low threshold that we choose for our rainy day cutoff. Indeed, those familiar with hiking in these mountains in the summer know that the frequency of the days with a thundershower somewhere within visible range is frequent and can go on for long periods.

The summer period (June, July, and August) dry spell analysis is shown in Figures 44 and 45. In these figures, we establish the cause of the strong model bias that results in nearly twice too much precipitation in all regions, and a strong summer heavy precipitation bias. In those regions where no strong summer bias was observed (Washington, Washington-Idaho, Idaho-Montana, Oregon, and Idaho), the wet spell lengths and distributions fall right on the observed distributions, but the dry spell lengths are considerably shorter in all of the areas. While the reduction in dry spell periods in the north and west is not too drastic, the length of the wet spell increases, shown in the previous figure, match the reduction in the dry spell periods for the south and west. Not only are the wet spells too long, but they are coming about twice too often.

III.4 DISCUSSION AND CONCLUSIONS

The regional model was found to be free of the grid-point storms of earlier versions. It also showed good skill at estimating precipitation associated with synoptic events. It did not perform as well in estimating sub-grid-scale convective precipitation, showing poor skill as compared to climatology for the Arizona mountains. The model also showed lower diurnal variation than the observed data with the T_{max} and T_{min} values 3 to 5 degrees Celsius lower and higher, respectively. The differences were so symmetric that the model T_{ave} was almost identical to the observed.

We have presented the results of a 14-year continuous regional model simulation, designed to evaluate the usefulness of such a system for studying GCM simulated climate change prognostications in the mountainous western US. We have compared our model results with a state-of-the-art observed data gridding process, not running on the complex actual terrain, but on the smoother model surface. We have also shown the need for much longer model runs in evaluating regional climate models. Model skill in each of our tests varied strongly with season and prior conditions, especially in the area precipitation fields. Clearly, randomly choosing a wet and dry year for comparison does not capture the full story of model performance. We now discuss some of the implications of our comparison.

In this comparison, we apply two approaches to providing climate data inputs to an artificial surface. The reduced, diurnal temperature swing produced by the model, and its twice too much recorded rainfall, is easily judged unreasonable. If our statistical tools are to be of value, they should help us understand the mechanisms leading to these unreasonable values. On the other hand, the limited resolution and sampling of the observed network leaves us to wonder about the limits of representativeness of these estimates for some of the sub-grid-scale processes, such as mountain peak summer afternoon rain showers. This is verified by the fact that observed error levels for the daily data are even larger than the model bias.

The model skill in estimating temperatures, once the lower than observed diurnal bias is removed, is quite good. The general underestimation of the diurnal temperature range in the model output has been seen in other simulations [*Giorgi et al.*, 1993a]. We have investigated whether the

reduced diurnal variation is a model problem, or related to its surface treatment model (BATS). BATS sensitivity and quality have been examined closely by several reviewers [*Giorgi et al.*, 1994; *Xiaogang et al.*, 1996; *Seth and Giorgi*, 1996]. Since the model's lowest level is at about 20 m above the model surface, it should have a lower daily variation than the observed temperatures that were collected on 1.2 – 1.5 m above the surface. Our use of shelter temperatures in this analysis was an attempt to correct this problem. The shelter temperature routine as outlined in chapter II uses an adjustment between the skin temperature and the lowest model layer. Recently, *Jin et al.* [1997] published a similar comparison between the BATS-1E skin temperature derived from CCM2 runs and a global satellite derived skin temperature and found a "useful" agreement. However, our skin temperature diurnal variation is smaller than the observed screen height maximum – minimum temperature difference. As a result, the change to a surface air temperature did not improve the maximum and minimum temperature biases. We intend to continue to examine this discrepancy in a future effort.

Adjustments in the precipitation modeling scheme appear to have solved earlier problems with large grid-point storms, but have not kept the summer time precipitation from being overestimated in these mountain regions. In other areas not associated with mountains, the model actually underestimates precipitation [*Giorgi et al.*, 1994]. In making our examination of the statistics of the precipitation pattern, the probability mass functions of both the intensity plots and the spell analysis series point back to a problem with these models just precipitating too much. One possible change to the model to compensate for this problem is to use explicit cloud and rain water physics. Although this is unlikely to change the poor performance of the convective precipitation, explicit cloud and rain water may improve grid-scale resolvable precipitation.

An advantage of using the explicit cloud and rain water parameterization is that it has tuning parameters that could be adjusted to improve the resultant simulation. The poor performance of the convective parameterization is a known problem, with many different attempts at providing better solutions. It is unlikely that simple adjustments to the parameterization might yield better results. The main problem is that storms are triggered too often. This can be because the potential temperature profiles indicate instabilities too frequently or the large-scale triggering mechanisms in the convective parameterization are not functioning correctly. Perhaps too much water is being transported from the LBC and/or the Pacific Ocean. Analysis of the problems with convective precipitation, much less solutions, will be quite difficult.

The climatology of Washington-Idaho and Oregon suggests that the model is too wet throughout the year by 50 – 100 mm/month, with the biggest differences in winter. In the other regions the problem is predominantly with summer precipitation. Clearly, the model is still struggling with the prediction of summertime convective precipitation in high mountain regions. The precipitation threat score plots (Figures 29 and 30) tell the story best. The model has high skill during the winter and spring months and much lower scores during the summer and fall periods. It also simulates the pattern of precipitation (wet – dry days) pretty well, except the wet spells in summer are too long as currently calculated. As precipitation amount increases, accuracy decreases. This is a typical problem with weather simulation models and accounts for the forecasters often reporting that there is the chance of rain but not the amounts. The fact that the model does as well as it does in this stressful situation is quite impressive, considering that real data is up to a 1000 km away and the model has full freedom to swish things around in the middle. Examining only 250 x 250 km boxes lowers our scores, because the model can move a storm through the system on a slightly

different path than that actually observed. When closely examined, the seasonality in the skill plots is not surprising. The model's sub-grid-scale component of rain can be considered a random signal that is not easily predicted. When the model is handling a structured synoptic event, such as a front, the code works well. However, when the model develops conditions for localized thunderstorms to develop, it simply does not have enough information on where to let those cells develop. A clever approach to this problem was demonstrated by *Leung and Ghan* [1994]. They provided their model with sub-grid-scale terrain and land use information as well as a parameterization scheme and used this information in the precipitation generation routines. This allows a model to consider sub-grid effects without having to do all the calculations associated with the higher scale information.

Now we come back to the original objective; can nested models be used at this point to downscale GCM simulations and to drive distributed ecosystem and hydrologic models in western US mountain terrain. *Limaye et al.* [1996; *Limaye*, 1998] used the output of our model study to drive a hydrologic model for a specific basin in northern Utah. The results were mixed. The first half of the simulation was modeled very well, but in the second half streamflow rates were much too high. In the beginning the model captured both the high and low streamflow years, but later on the regional model simply had too much precipitation and the hydrologic model consistently overpredicted streamflow rates. Since the simulation performed well for the first several years, there obviously is potential in this method. However, regional models need to be tuned based on longer integrations such as this study, rather than one or two simulation years. The regional model output needs to be rigorously compared to observations so the biases are understood before sending the output to other models. For the GCM downscaling cases, where for future climates the statistics of our current climate systems may be disrupted, the use of regional models may be appropriate. However, the biases and accuracy to existing climate (such as provided here) must also be considered in any conclusions.

References

Adams, D. K., and A. C. Comrie, The North American monsoon. *Bull. Am. Meteorol. Soc.*, *78, 10*, 2197-2213, 1997.

Anthes, R. A., Hsie E.-Y., Kuo Y.-H., Description of the Penn State/NCAR mesoscale model version 4 (MM4), *NCAR Tech. Note, TN-282+STR*, 66 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1987.

Arakawa A., and W. H. Shubert, Interaction of a cumulus cloud ensemble with the large scale environment. Part I, *J. Atmos. Sci.*, *31*, 674-701, 1974.

Avissar, R., and R. A. Pielke, A parameterization of heterogenous land surfaces for atmospheric numerical models and its impact on regional meteorology. *Mon. Weather Rev.*, AMS, *117*, 2113-2136, 1989.

Bardossy, A., Downscaling from GCMs to local climate through stochastic linkages, in *Climate Change, Uncertainty and Decision-Making*, Published by The Institute for Risk Research, Ontario, Canada, pp. 33-46, 1994.

Barros A. P., and D. P. Lettenmaier, Dynamic modeling of the spatial distribution of precipitation in remote mountainous areas, *Mon. Weather Rev.*, *121*, 1195-1214, 1993.

Battisti, D. S., C. M. Bitz, and R. E. Moritz, Do general circulation models underestimate the natural variability in the arctic climate?, *J. Clim., 10*, 1909-1920, 1997.

Briegleb, B. P., Delta-Eddington approximation for solar radiation in the NCAR Community Climate Model, *J. Geophys. Res., 97*, 7603-7612, 1992.

Chen, T. H., A. Henderson-Sellers, P. C. D. Milly, A. J. Pitman, A. C. M. Beljaars, J. Polcher, F. Abramopoulos, A. Boone, S. Chang, F. Chen, Y. Dai, C. E. Desborough, R. E. Dickinson, L. Dumenil, M. Ek, J. R. Garratt, N. Gedney, Y. M. Gusev, J. Kim, R. Koster, E. A. Kowalczyk, K. Laval, J. Lean, D. Lettenmaier, X. Liang, J. -F. Mahfouf, H. -T. Mengelkamp, K. Mitchell, O. N. Nasonova, J. Noilhan, A. Robock, C. Rosenzweig, J. Schaake, C. A. Schlosser, J. -P. Schulz, Y. Shao, A. B. Shmakin, D. L. Verseghy,P. Wetzel, E. F. Wood, Y. Xue, Z. -L. Yang, Q. Zeng, Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes, *J. Clim., 10*, 1194-1215, 1997.

Copeland, J. H., R. A. Pielke, and G. F. Kittel, Assessing climate and vegetation change for the central grasslands biogeographic region, in *Proceedings of the 6th Conference on Climate Variations*, AMS, Nashville, TN., 23-28, 1994.

Corby G. A., A. A. Gilchrist, and R. L. Newson, A general circulation model of the atmosphere suitable for long period simulations, *Q. J. R. Meteorol. Soc., 98*, 809-832, 1972.

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, Biosphere-atmosphere transfer scheme (BATS) version 1E as coupled to the NCAR community climate model, *NCAR/TN-387+STR*, 69 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1993.

Dodson, R., and D. Marks, Daily air temperature interpolated at high spatial resolution over a large mountainous region, *Clim. Res., 9*, 1-20, 1997.

Fulker, D., S. Bates, and C. Jacobs, Unidata: A virtual community sharing resources via technological infrastructure, *Bull. Am. Meteorol. Soc , 78, 3*, 457-468, 1997.

Gaffen, D. J., Evaluation of tropospheric water vapor simulations from the atmospheric model intercomparison project, *J. Clim., 10*, 1648, 1997.

Gaffen, D. J., E. W. Friday, Jr., D. E. Weston, D. B. Franklin, Infrared radiance spectra for testing radiative transfer models in cold and dry atmospheres: test cases from the Antarctic Plateau, *Bull. Am. Meteorol. Soc., 78, 10*, 2246-2247, 1997.

Giorgi, F., Simulation of regional climate using a limited area model nested in a general circulation model, *J. Clim., 3*, 941-963, 1990.

Giorgi, F., Sensitivity of simulated summertime precipitation over the western United States to different physics parameterizations, *Bull. Am. Meteorol. Soc., 119*, 2870-2888, 1991.

Giorgi, F., and G. T. Bates, The climatological skill of a regional model over complex terrain, *Mon. Weather Rev., 117*, 2325-2347, 1989.

Giorgi, F., G. T. Bates, and S. J. Nieman, Simulation of the arid climate of the Southern Great Basin using a regional climate model. *Bull. Amer. Meteorol. Soc., 73*, 1807-1822, 1992.

Giorgi, F., G. T. Bates, and S. J. Nieman, The multiyear surface climatology of a regional atmospheric model over the western United States, *J. Clim., 6*, 76-95, 1993a.

Giorgi, F., C. S. Brodeur, and G. T. Bates, Regional climate change scenarios over the united states produced with a nested regional climate model, *J. Clim., 7*, 375-399, 1994.

Giorgi, F., M. R. Marinucci, and G. T. Bates, Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes, *Mon. Weather Rev., 121*, 2794-2813, 1993b.

Giorgi, F., M. R. Marinucci, G. T. Bates, and G. D. Canio, Development of a second-generation regional climate model (RegCM2). Part II: Convective processes and assimilation of lateral boundary conditions, *Mon. Weather Rev., 121*, 2814-2832, 1993c.

Grell, G. A., J. Dudhia, and D. R. Stauffer, A description of the fifth generation Penn State/NCAR mesoscale model (MM5), NCAR Tech. Note, NCAR/TN-398+STR, 121pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1994.

Holtslag, A. A. M., and B. A. Boville, Local versus non local boundary-layer diffusion in a global climate model, *J. Clim., 6*, 1825-1842, 1992.

Holtslag, A. A. M., E. I. F. de Bruijin, and H. L. Pan, A high-resolution airmass transformation model for short-range weather forecasting, *Mon. Weather Rev., 118*, 1561-1575, 1990.

Hong, S.-Y., and H.-M. H. Juang, Orography blending in the lateral boundary of a regional model, *Mon. Weather Rev., 126*, 1714-1718, 1998.

Hostetler, S. W., and F. Giorgi, Use of a regional atmospheric model to simulate lake-atmosphere feedbacks associated with Pleistocene lakes Lahontan and Bonneville, *Clim. Dyn., 7*, 39-44, 1992.

Jin, M., R. E. Dickinson, and A. M. Vogelmann, A comparison of CCM2-BATS skin temperature and surface-air temperature with satellite and surface observations, *J. Clim., 10*, 1505-1524, 1997.

Leung, L. R., and S. J. Ghan, A subgrid parameterization of orographic precipitation, *Theor. Appl. Clim., 52*, 95-118, 1995.

Limaye, A. S., Development of hydrologic response unit delineation criterion and atmospheric hydrologic model link, Ph.D. dissertation, Utah State University, Logan, Utah, 104 pp., 1998.

Limaye A. S., E. B. Kluzek, G. E. Bingham, J. P. Riley, Linking Atmospheric and Hydrologic Models at the Basin Scale, *Phys. Chem. Earth, 21, 3*, 211-218, 1996.

Luthi, D., A. Cress, H. C. Davies, C. Frei, and C. Schar, Interannual variability and regional climate simulations, *Theor. Appl. Clim., 53*, 185-209, 1996.

Madala, R. V., Efficient time integration schemes for atmosphere and ocean models, in *Infinite-Difference Techniques for Vectorized Fluid Dynamics Calculations*, edited by D. L. Book, pp. 56-74, Springer-Verlag Publ., New York, 1981.

Matthews, D. M., F. Giorgi, and G. T. Bates, Regional model simulation of an extreme precipitation event in the Colorado river basin, in *Proceedings of the Third Symposium on Global Change*, AMS, Boston, Mass., 20-23, 1992.

Medina, J. G., Application of a simple local-scale numerical model in the study of altered climate impacts on watershed precipitation, in *Proceedings of the Second Symposium on Global Change Studies*, AMS, New Orleans, La, 68-73, 1991.

Min, W., The climate signal in regional moisture fluxes: A comparison of three global data assimilation products, *J. Clim.*, *10*, 2623-2642, 1997.

Oborn, S. M., G. Bingham, V. A. Nuthulapaty, E. B. Kluzek, and D. Partridge, A video display system for quality control of hydro-meteorological model data, paper presented at Fall meeting of the AGU, San Fransisco, Calif., Dec 5-9, 1994.

Paoli, G. M., *Climate Change, Uncertainty and Decision-Making*, 164 pp., Institute for Risk Research, University of Waterloo, Waterloo Ontario, Canada, 1994.

Palutikof, J. P., J. A. Winkler, C. M. Goodess, and J. A. Andresen, The simulation of daily temperature time series from GCM output. Part I: Comparison of model data with observations, *J. Clim.*, *10*, 2497-2513, 1997.

Portman, D. A., W. Wang, and T. R. Karl, Comparison of general circulation model and observed regional climates: Daily and seasonal variability, *J. Clim.*, *5*, 343-353, 1992.

Privalsky, V., T. E. Croley II, Statistical validation of GCM-simulated climates for the U.S. Great Lakes and the C.I.S. Emba and Ural river basins, *Stochastic Hydrol. Hydraul.*, *6*, 69-80, 1992.

Rajagopalan, B., U. Lall, and D. G. Tarboton, Evaluation of kernel density estimation methods for daily precipitation resampling, *Stochastic Hydrol. Hydraul.*, *11*, 523-547, 1997.

Reynolds, R. W., A real-time global sea surface temperature analysis, *J. Clim.*, *1*, 75-86, 1988.

Reynolds, R. W., and D. C. Marsico, An improved real-time global sea surface temperature analysis, *J. Clim.*, *6*, 114-119, 1993.

Seth, A., and F. Giorgi, Three-dimensional model study of organized mesoscale circulations induced by vegetation, *J. Geophys. Res.*, *101*, *D3*, 7371-7391, 1996.

Trenberth, K. E., and J. G. Olson, ECMWF global analyses 1979-1986: Circulation statistics and data evaluation, *NCAR Tech. Note.*, *NCAR/TN-300+STR*, 94 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1988.

Turner, D. P., R. Dodson, and D. Marks, Comparison of alternative spatial resolutions in the application of a spatially distributed biogeochemical model over complex terrain, *Ecol. Model.*, *90*, 53-67, 1996.

Wilson, L. L., and D. P. Lettenmaier, A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation, *J. Geophys. Res.*, *97*, *D3*, 2791-2809, 1992.

Woolhiser, D. A., and J. Roldan, Seasonal and regional variability of parameters for stochastic daily precipitation models, *Water Resour. Res.*, *22, 6*, 965-978, 1982.

Xiaogang, G., S. Sorooshian, and H. V. Gupta, Sensitivity analysis of the biosphere-atmosphere transfer scheme, *J. Geophys. Res.*, *101, D3*, 7279-7289, 1996.

Zhang, S., and E. Foufoula-Georiou, Subgrid-scale rainfall variability and its effects on atmospheric and surface variable predictions, *J. Geophys. Res.*, *102, D16*, 19559-19573, 1997.

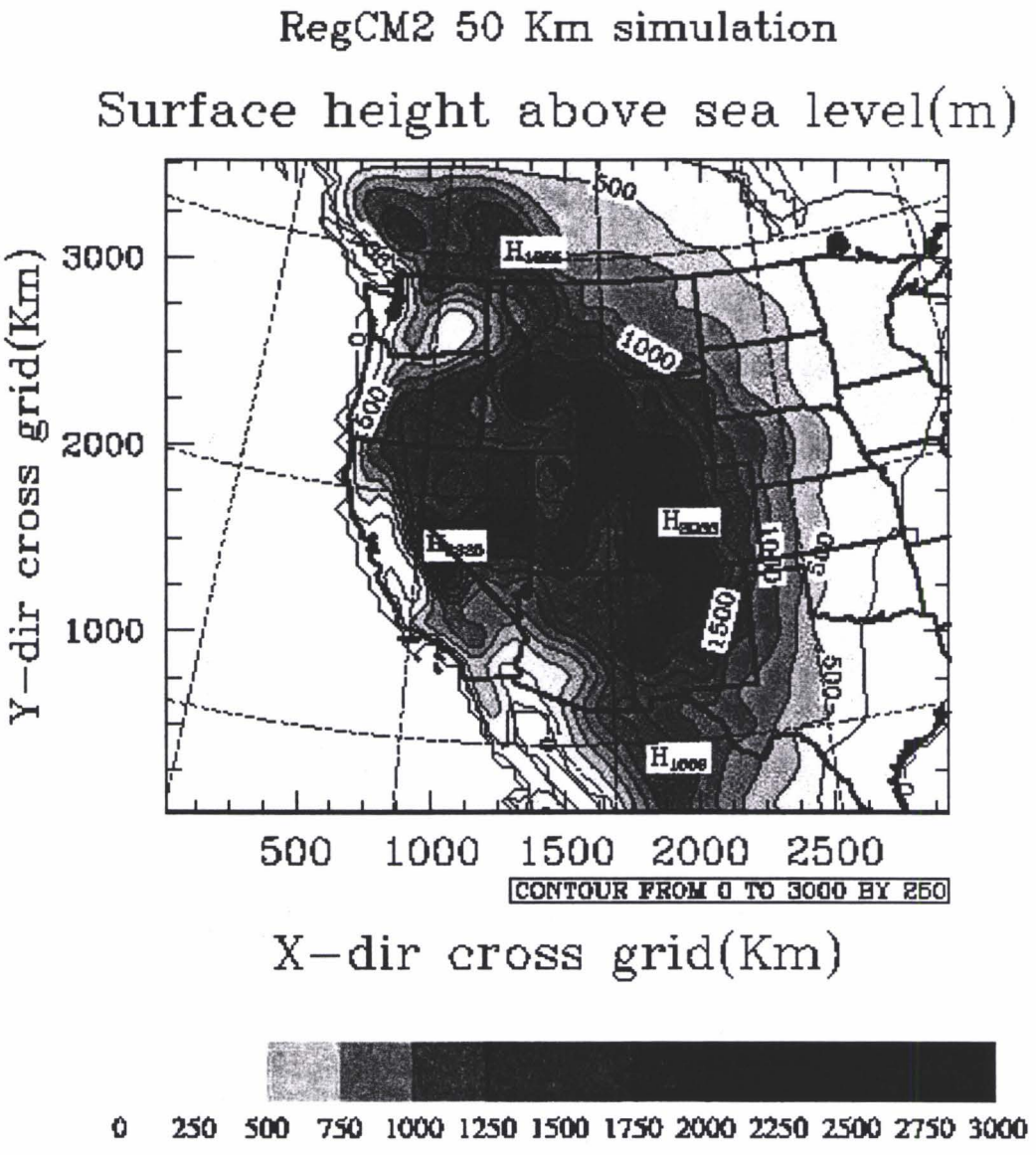

Figure 8. RegCM2 50 km terrain contour map of the western US. Compare these smoothed profiles with the DEM generated terrain in Figure 1

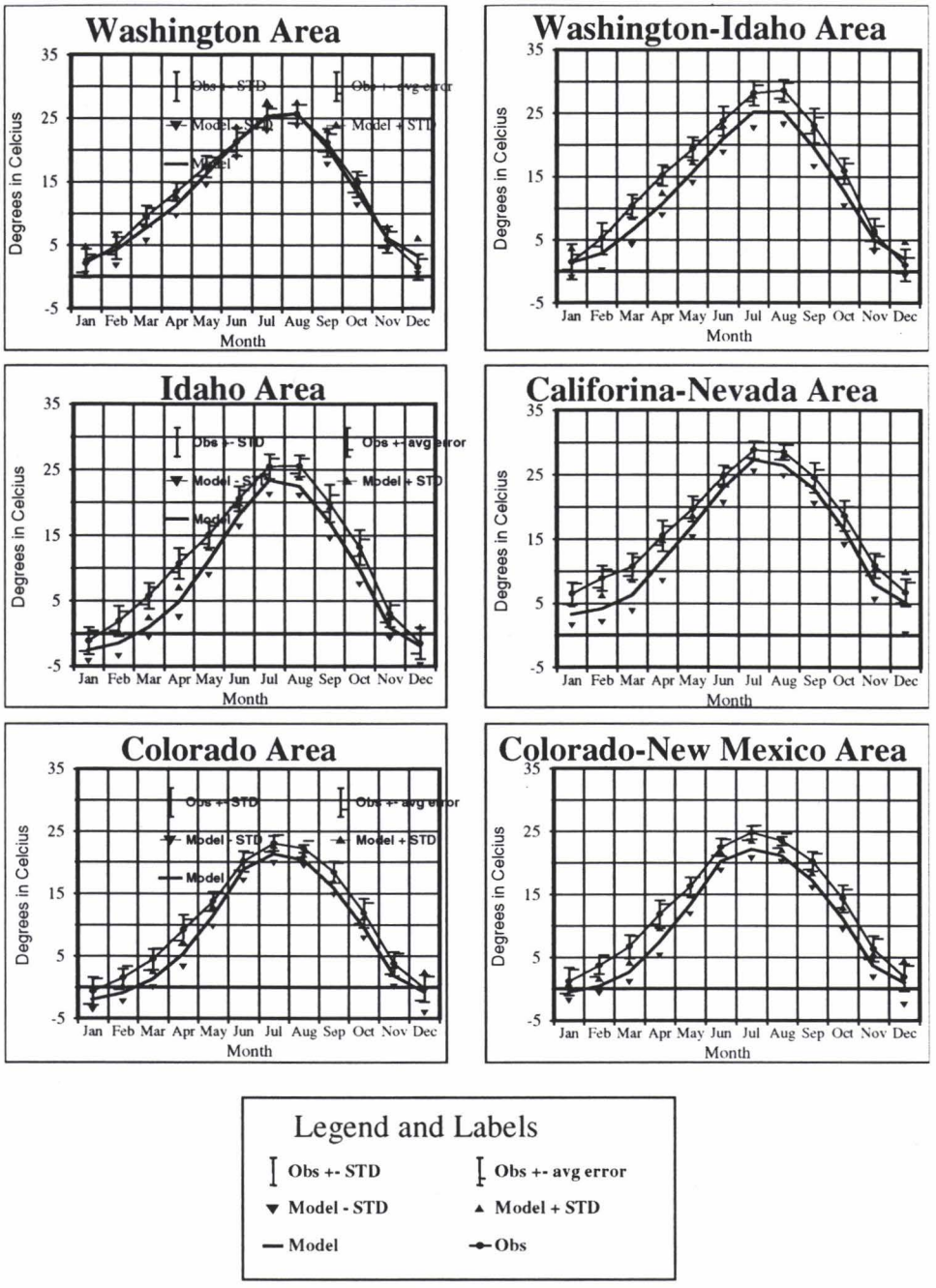

Figure 9. Maximum monthly temperature computed using the model and observed climatologies with standard deviations for six of the eleven analysis areas.

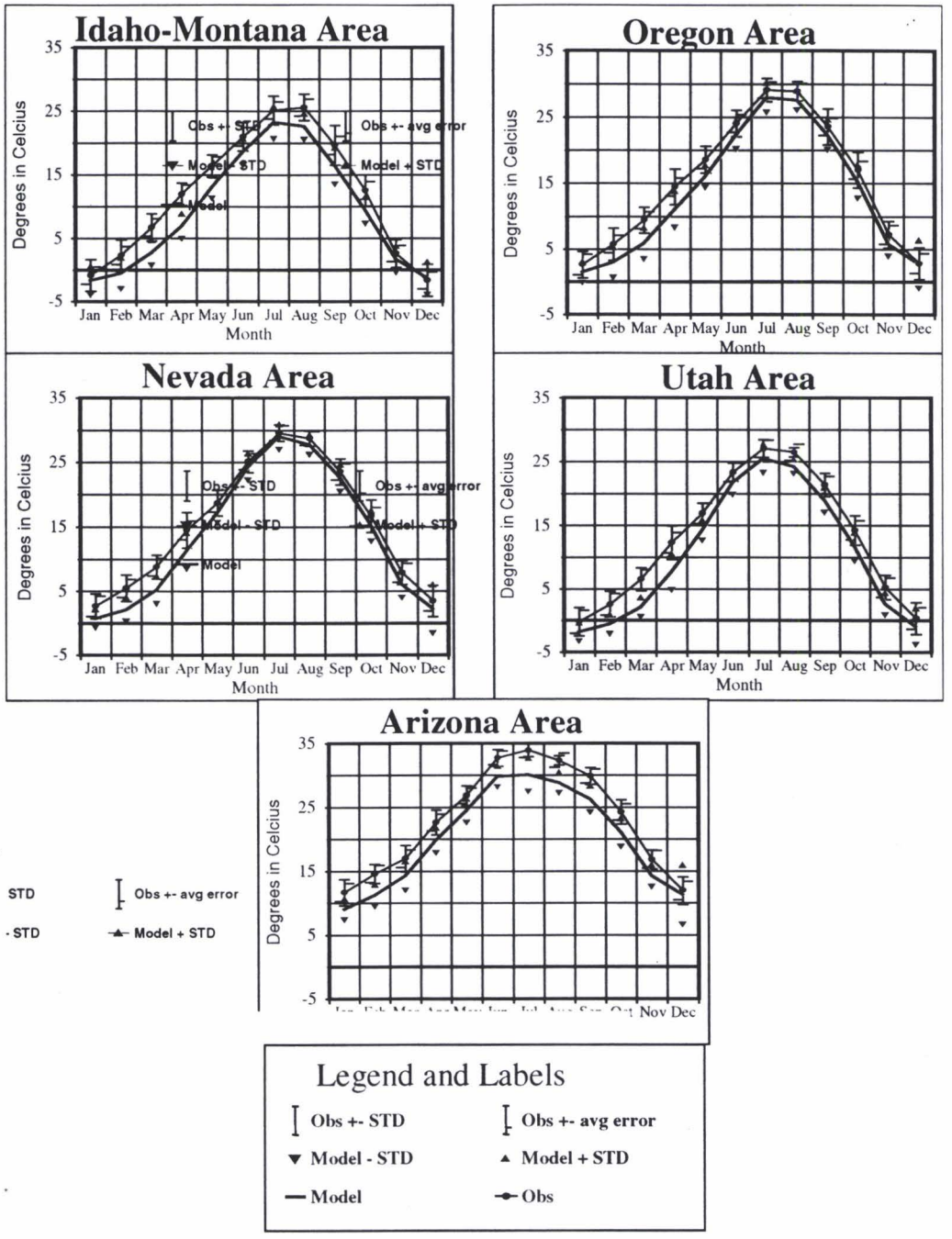

Figure 10. Maximum monthly temperature computed using the model and observed climatologies with standard deviations for five of the eleven analysis areas.

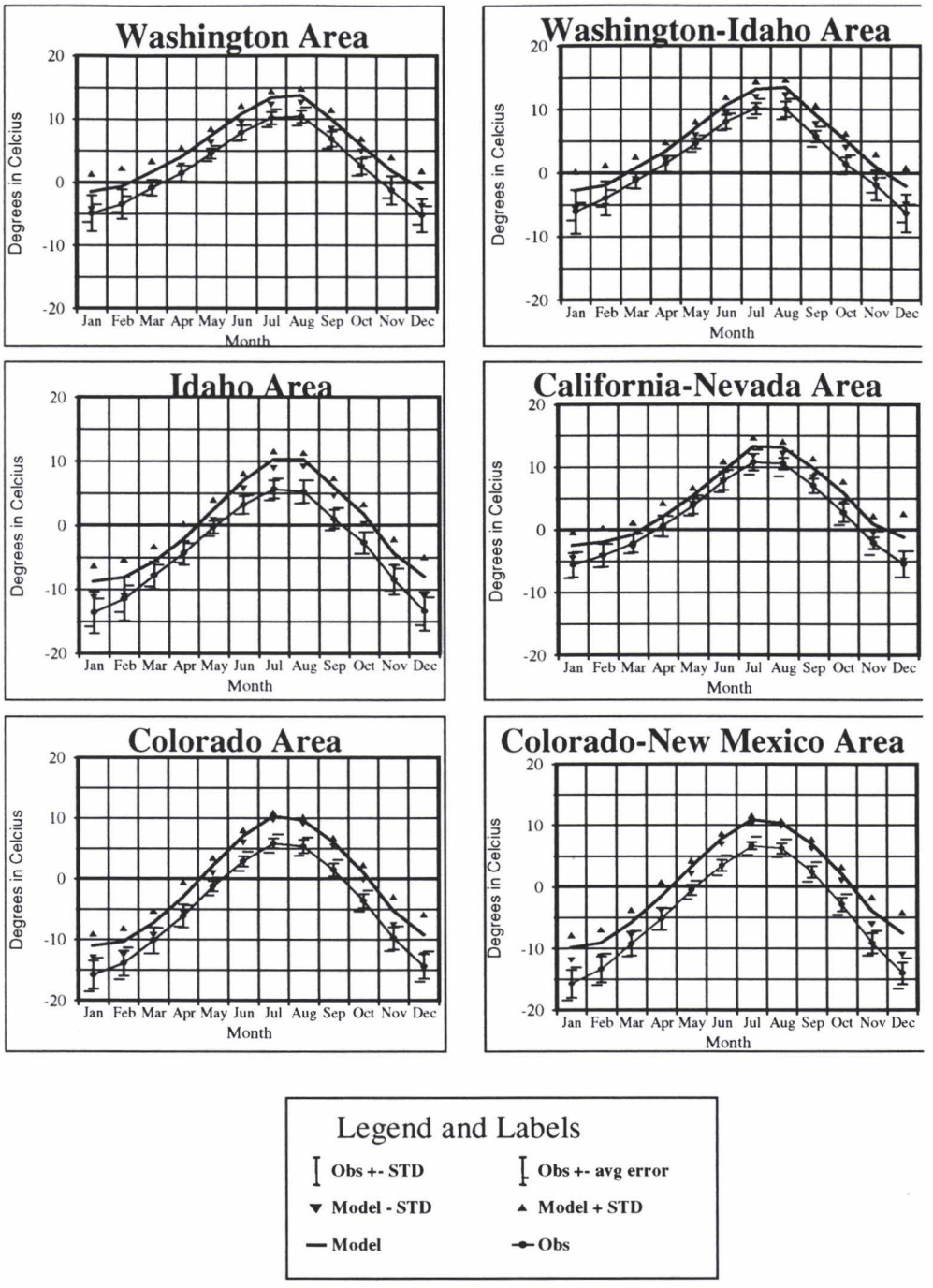

Figure 11. Plot of model and observed Minimum monthly temperature climatologies, with standard deviations for six of the eleven analysis areas.

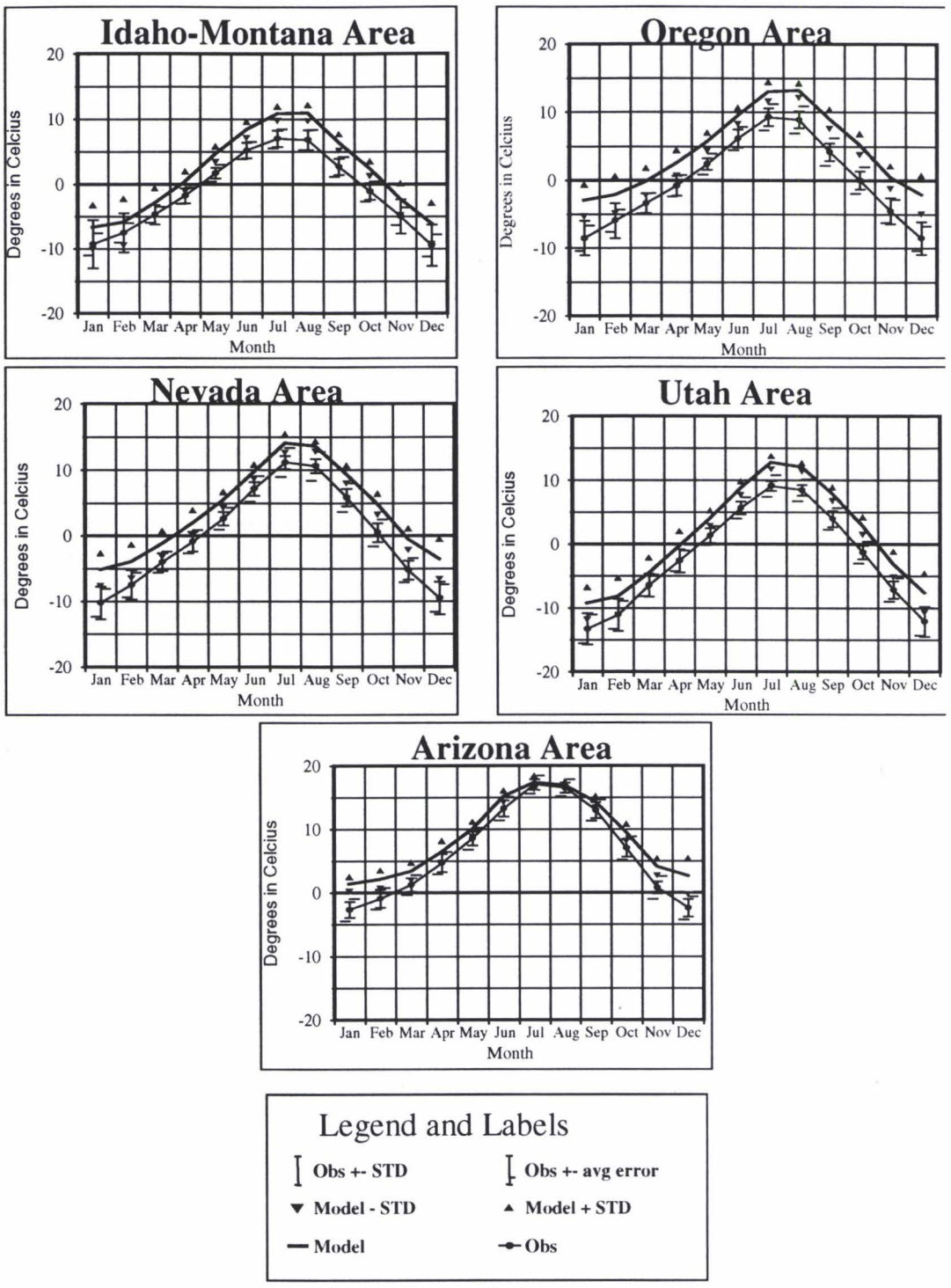

Figure 12. Plot of model and observed Minimum monthly temperature climatologies, with standard deviations for five of the eleven analysis areas.

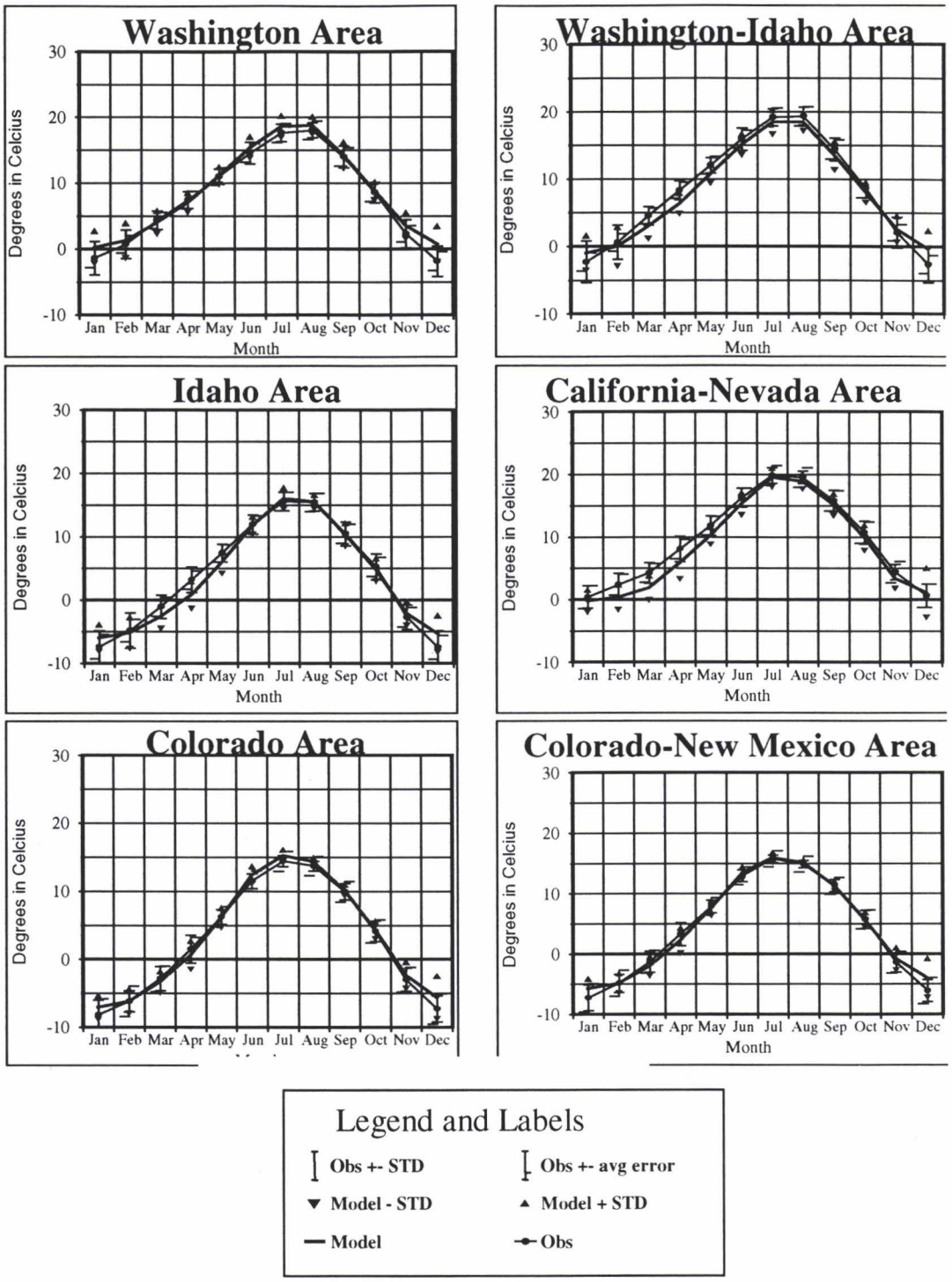

Figure 13. Mean monthly temperature computed using the model and observed climatologies with standard deviations for six of the eleven analysis areas.

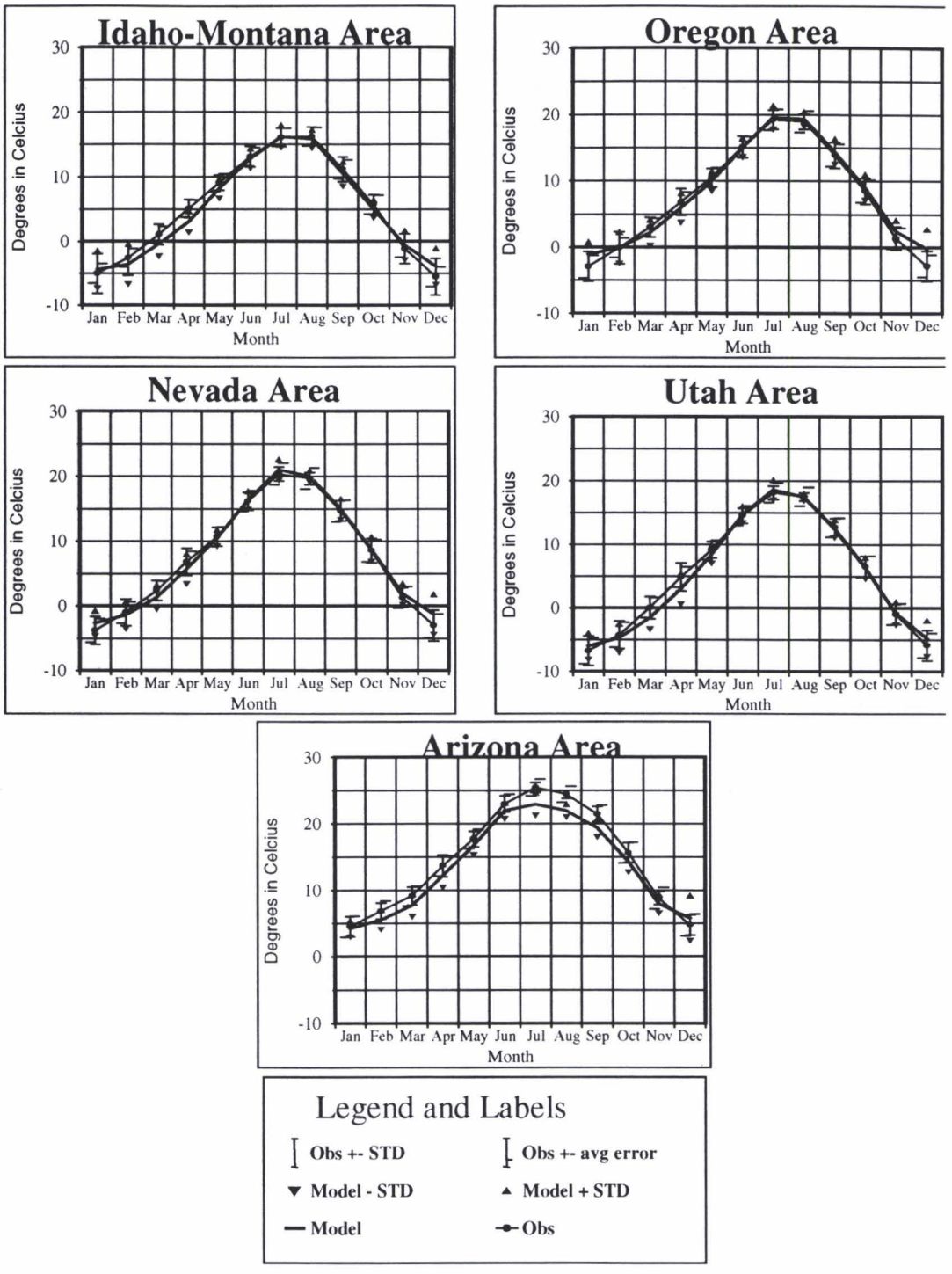

Figure 14. Mean monthly temperature computed using the model and observed climatologies with standard deviations for five of the eleven analysis areas.

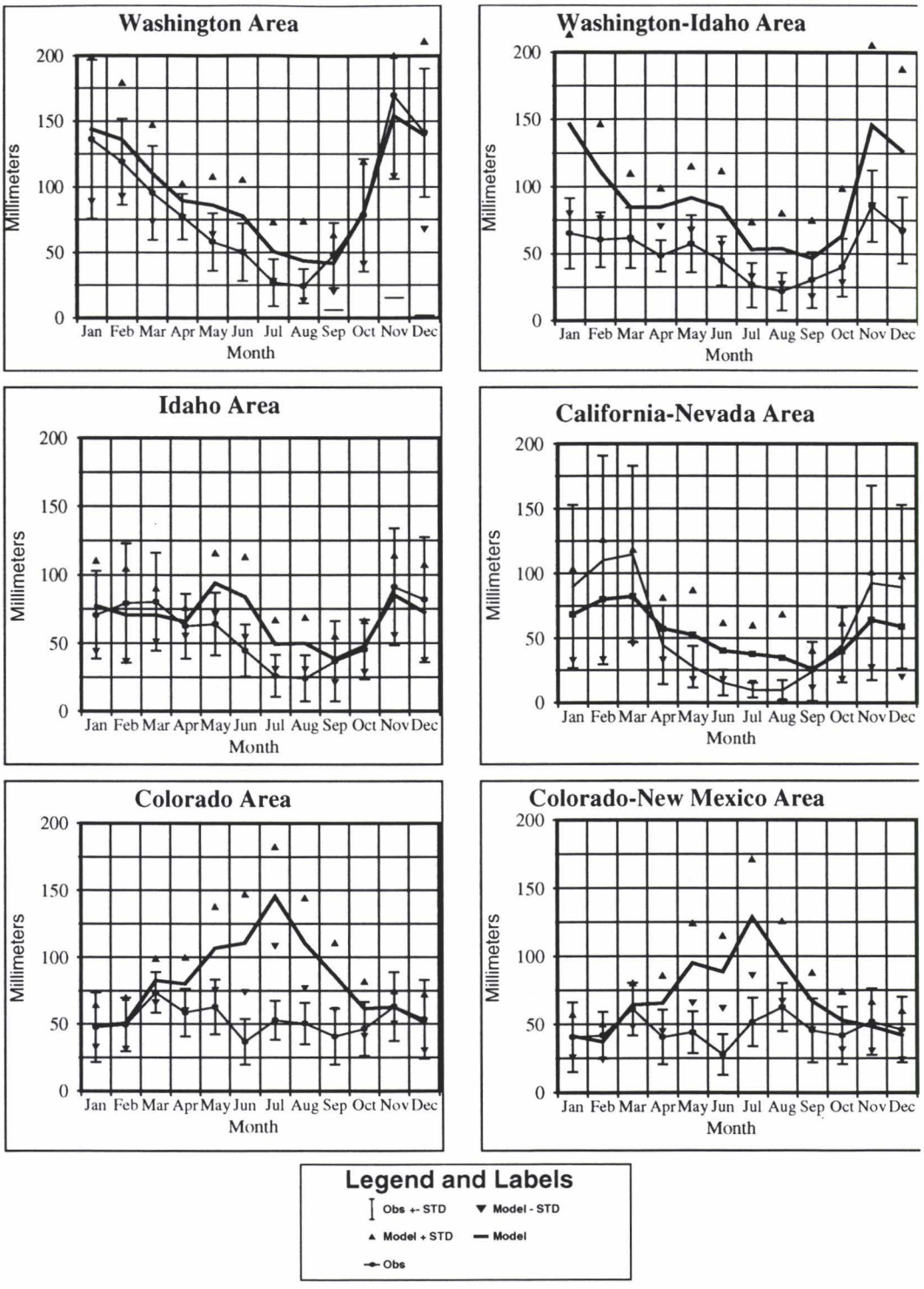

Figure 15. Monthly precipitation computed using the model and observed climatologies with standard deviations for six of the eleven analysis areas.

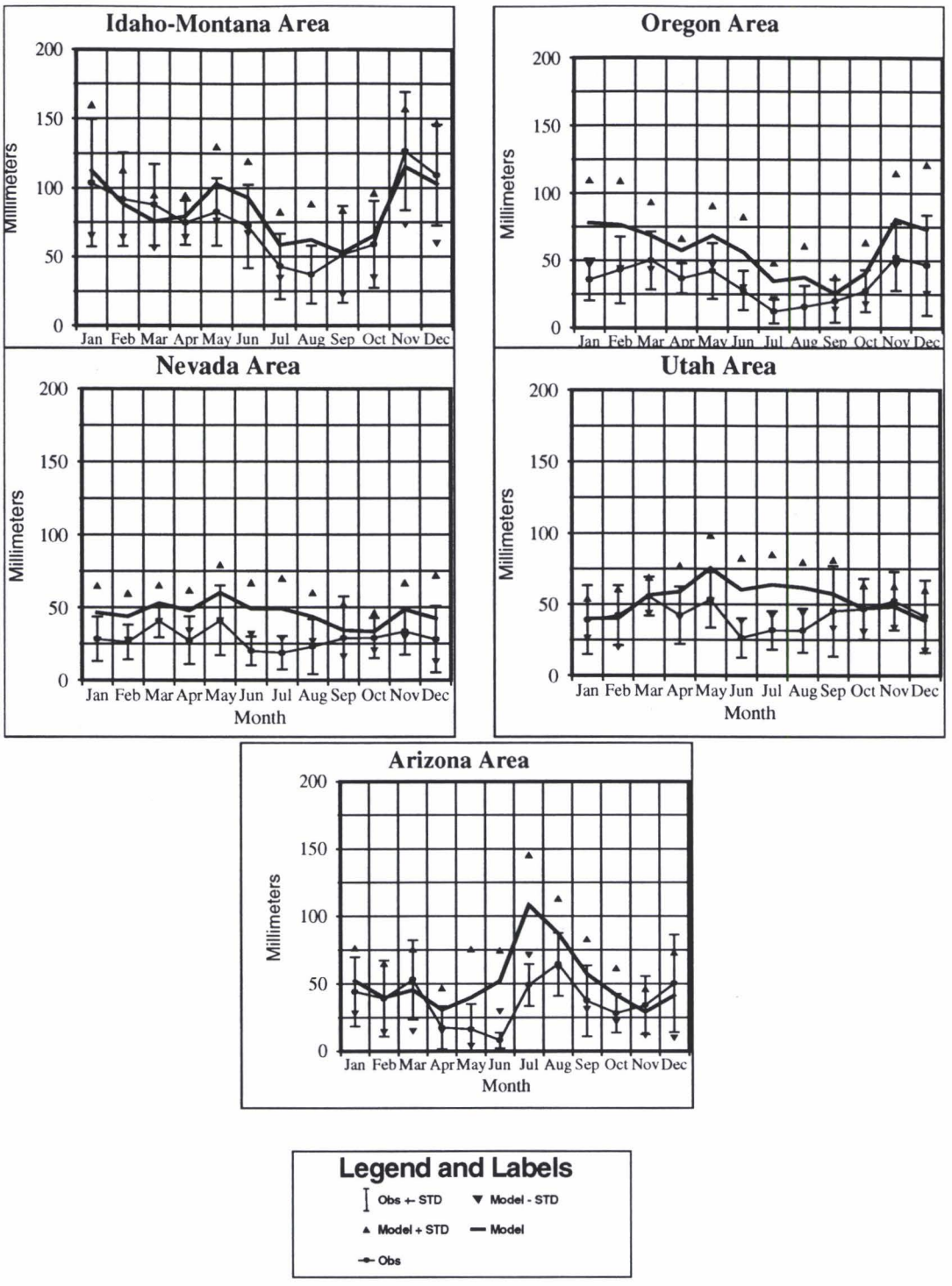

Figure 16. Monthly precipitation computed using the model and observed climatologies with standard deviations for five of the eleven analysis areas.

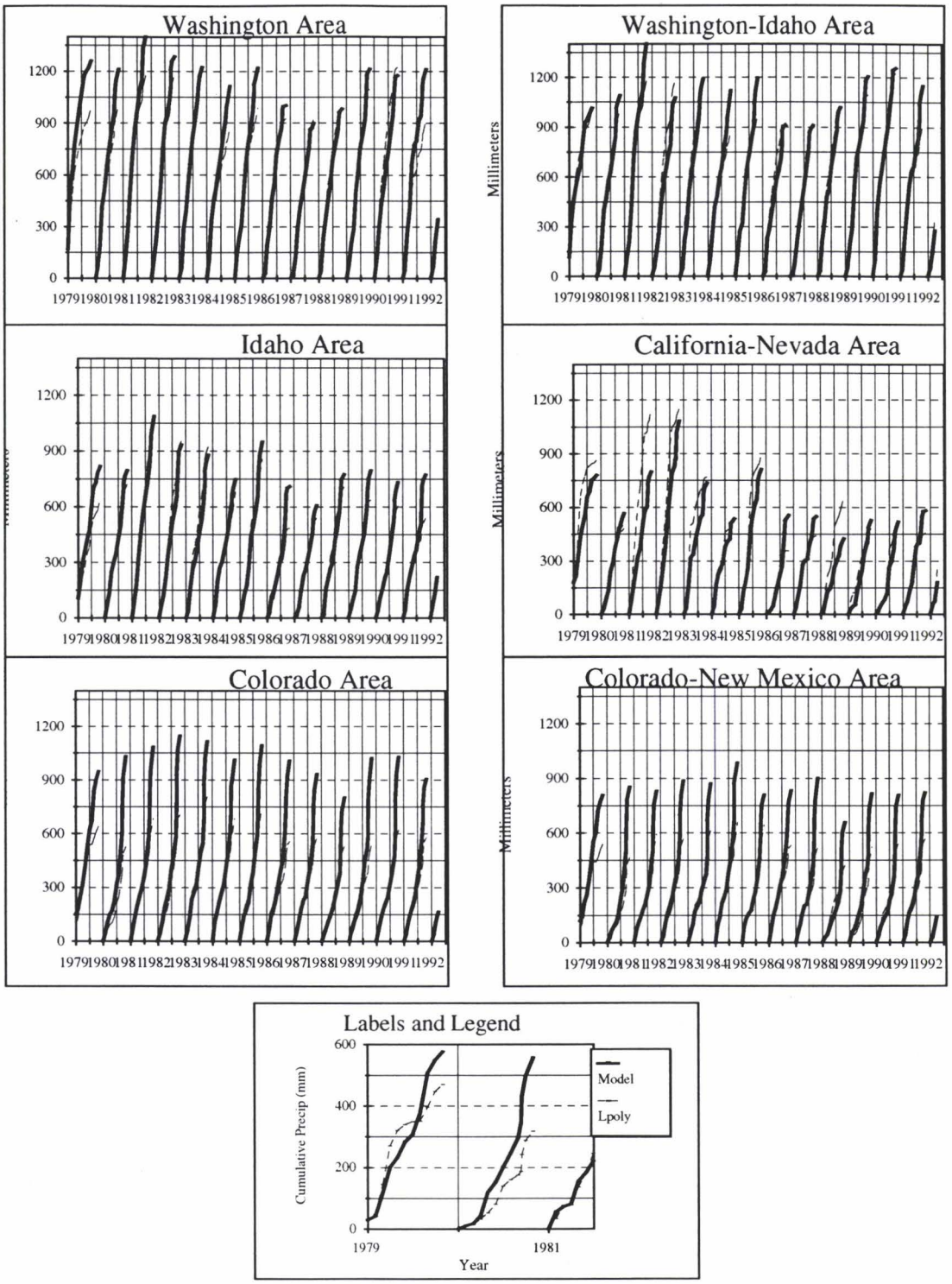

Figure 17. Cumulative monthly precipitation computed using the model and observed climatologies for six of the eleven analysis areas.

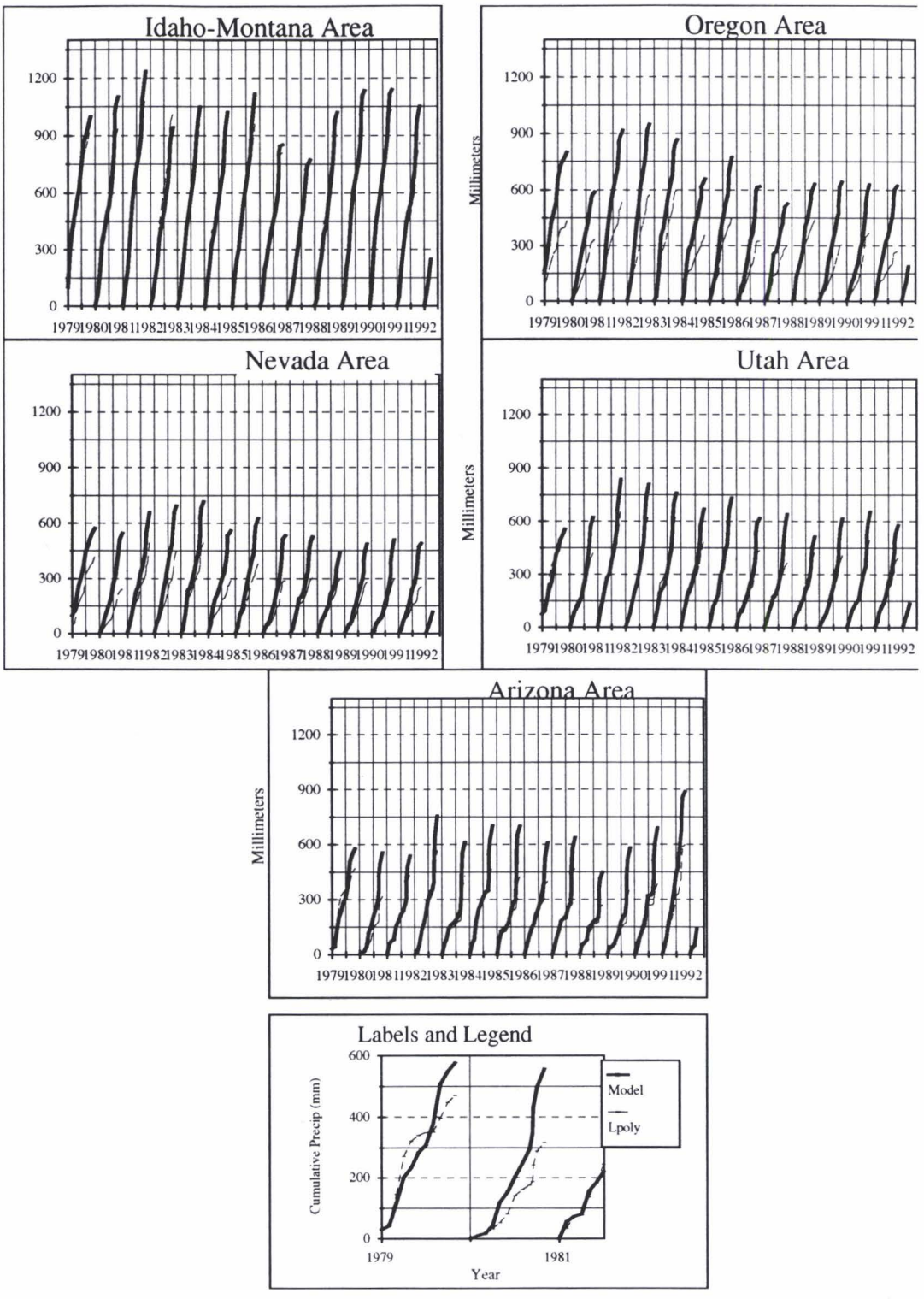

Figure 18. Cumulative monthly precipitation computed using the model and observed climatologies for five of the eleven analysis areas.

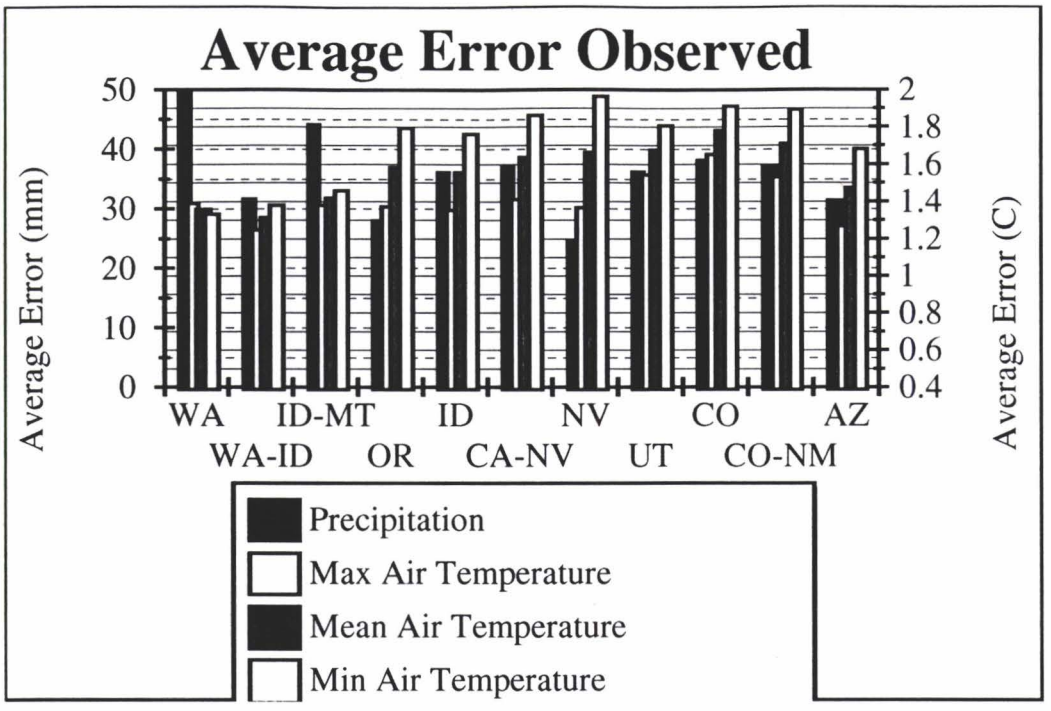

Figure 19. Annual average observed error for each region.

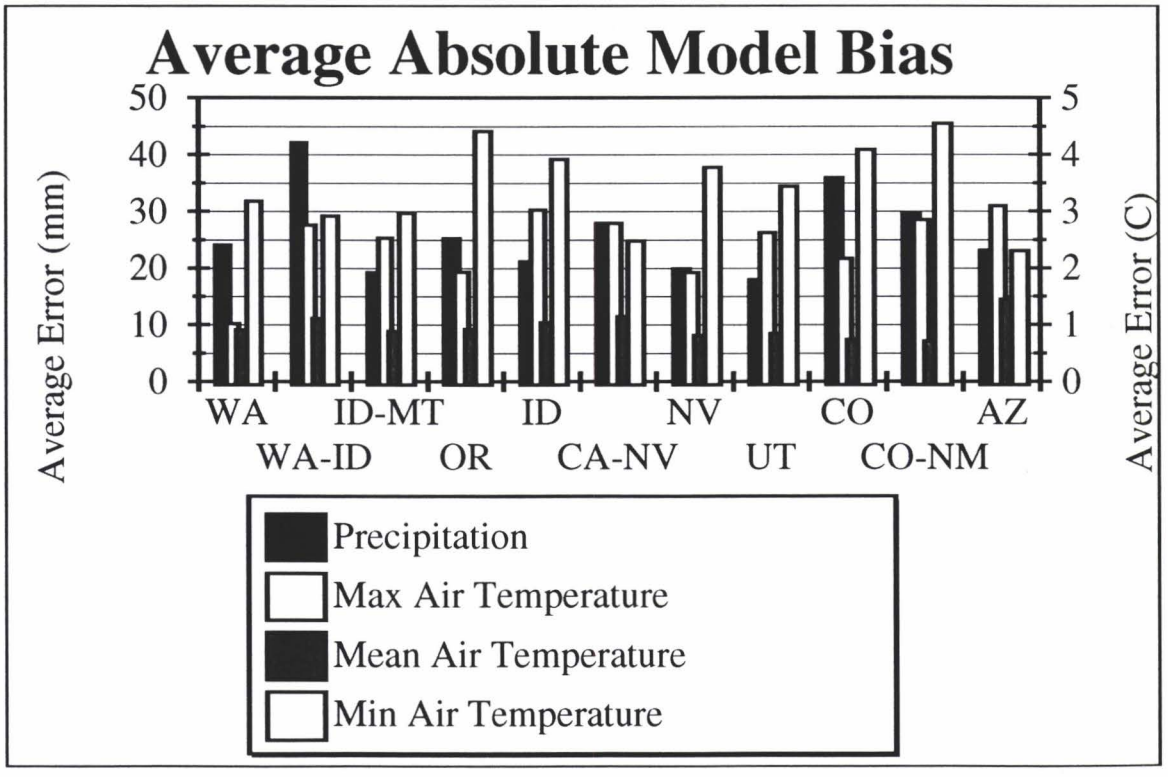

Figure 20. Model absolute bias between the model and observed climatologies.

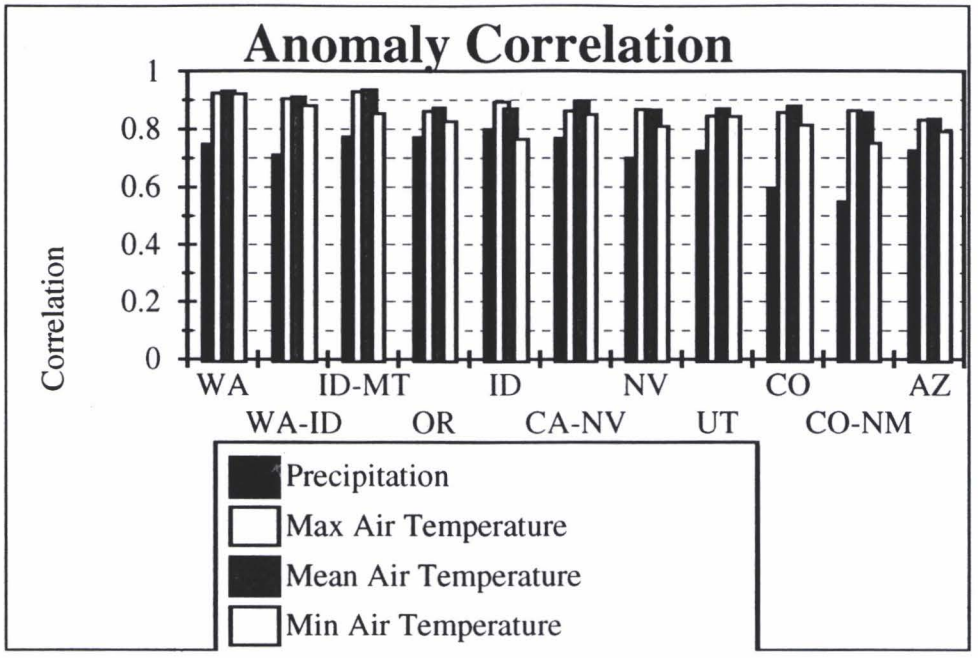

Figure 21. Correlation between the model and observed monthly anomalies.

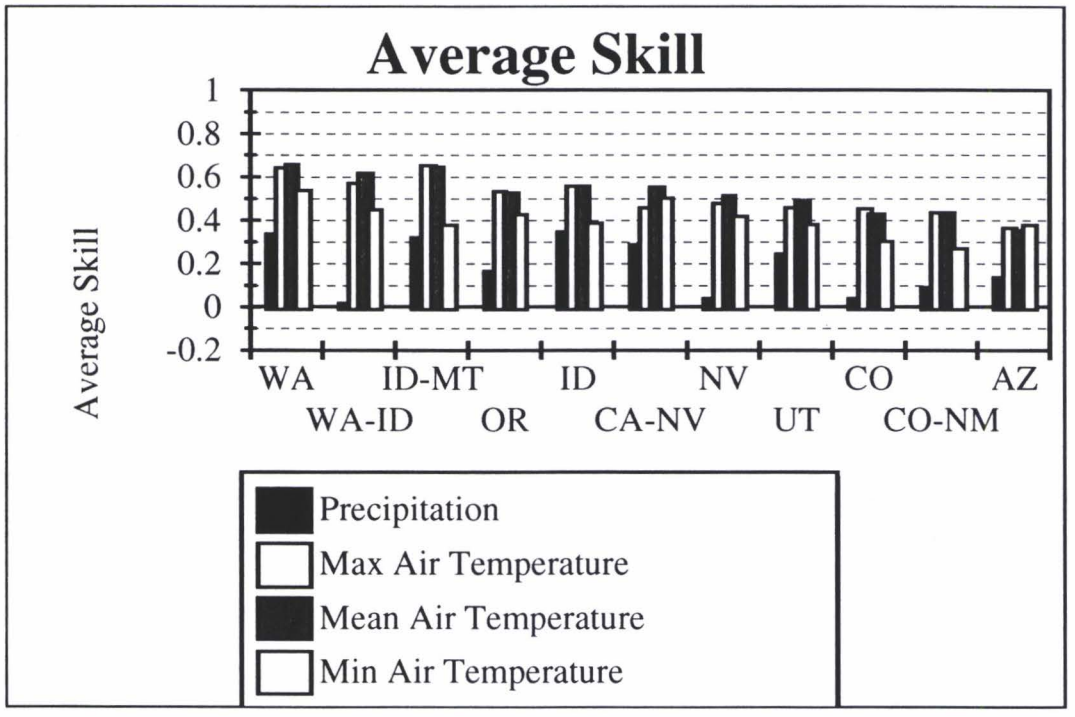

Figure 22. Model predictive skill compared with a predictive model based only on the observed climatology.

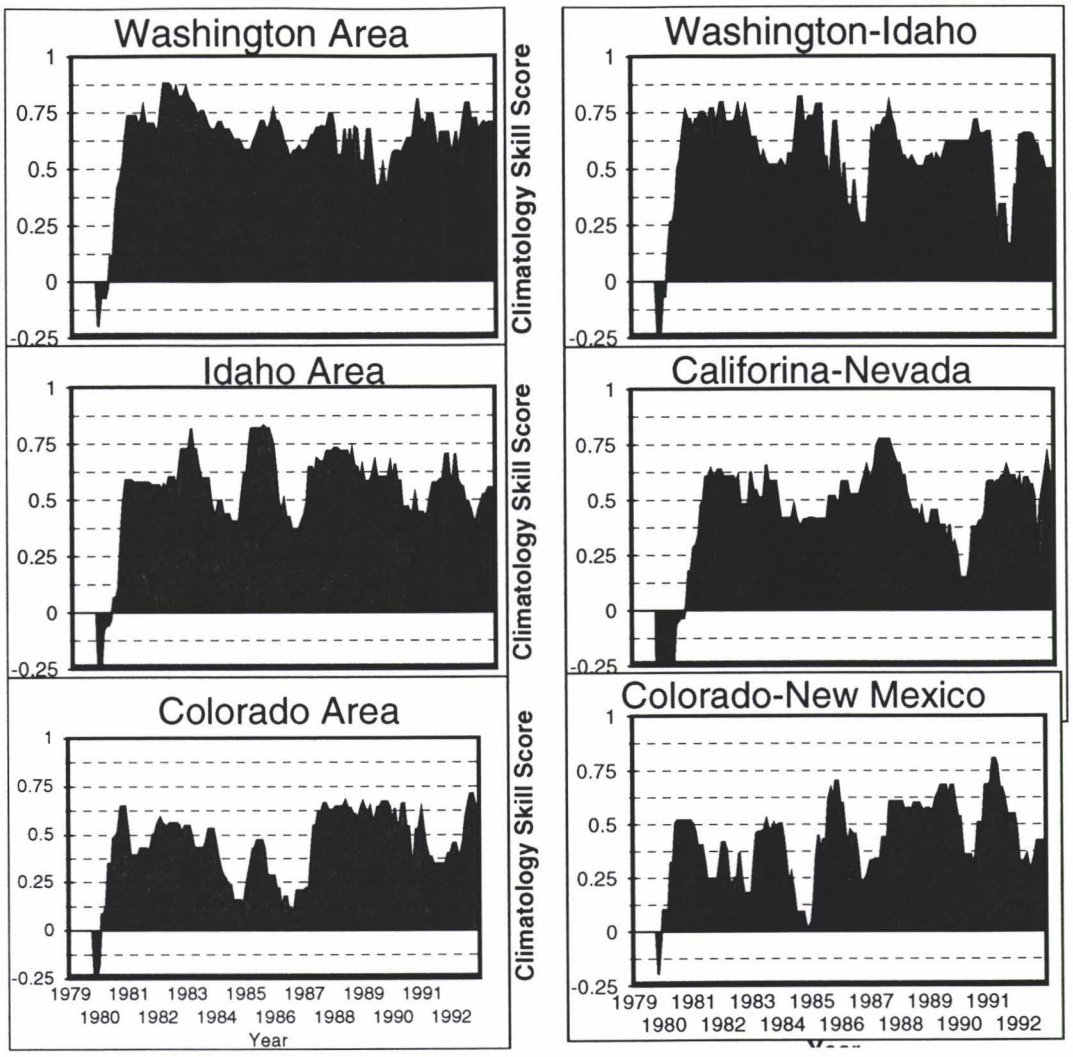

Figure 23. . Maximum air temperature model skill scores developed using a simple model based on the bias corrected observed climatology (perfect skill = 1). These plots are smoothed using a 30-day running mean.

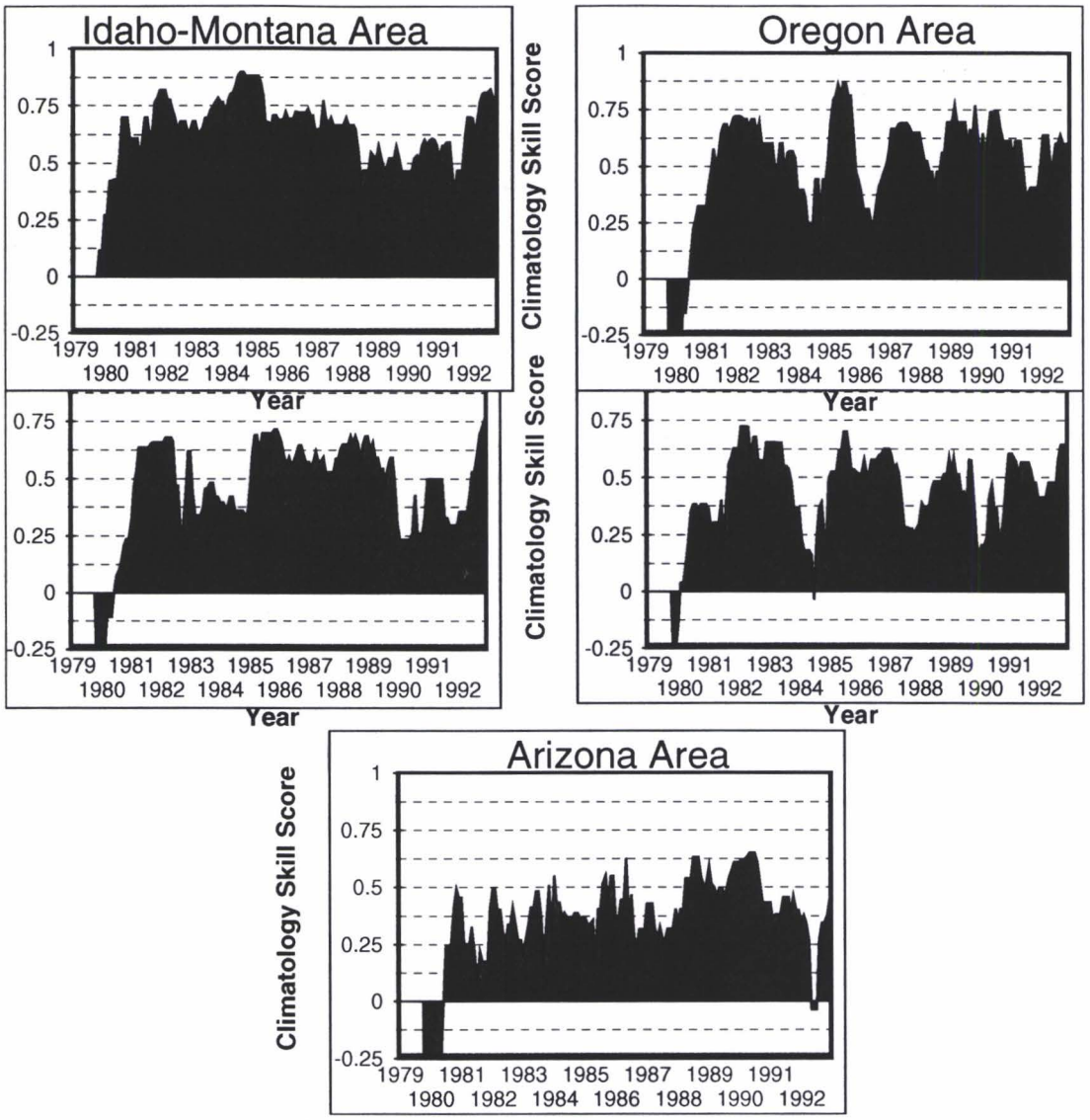

Figure 24. Maximum air temperature model skill scores developed using a simple model based on the bias corrected observed climatology (perfect skill = 1). The plots are smoothed using a 30-day running mean.

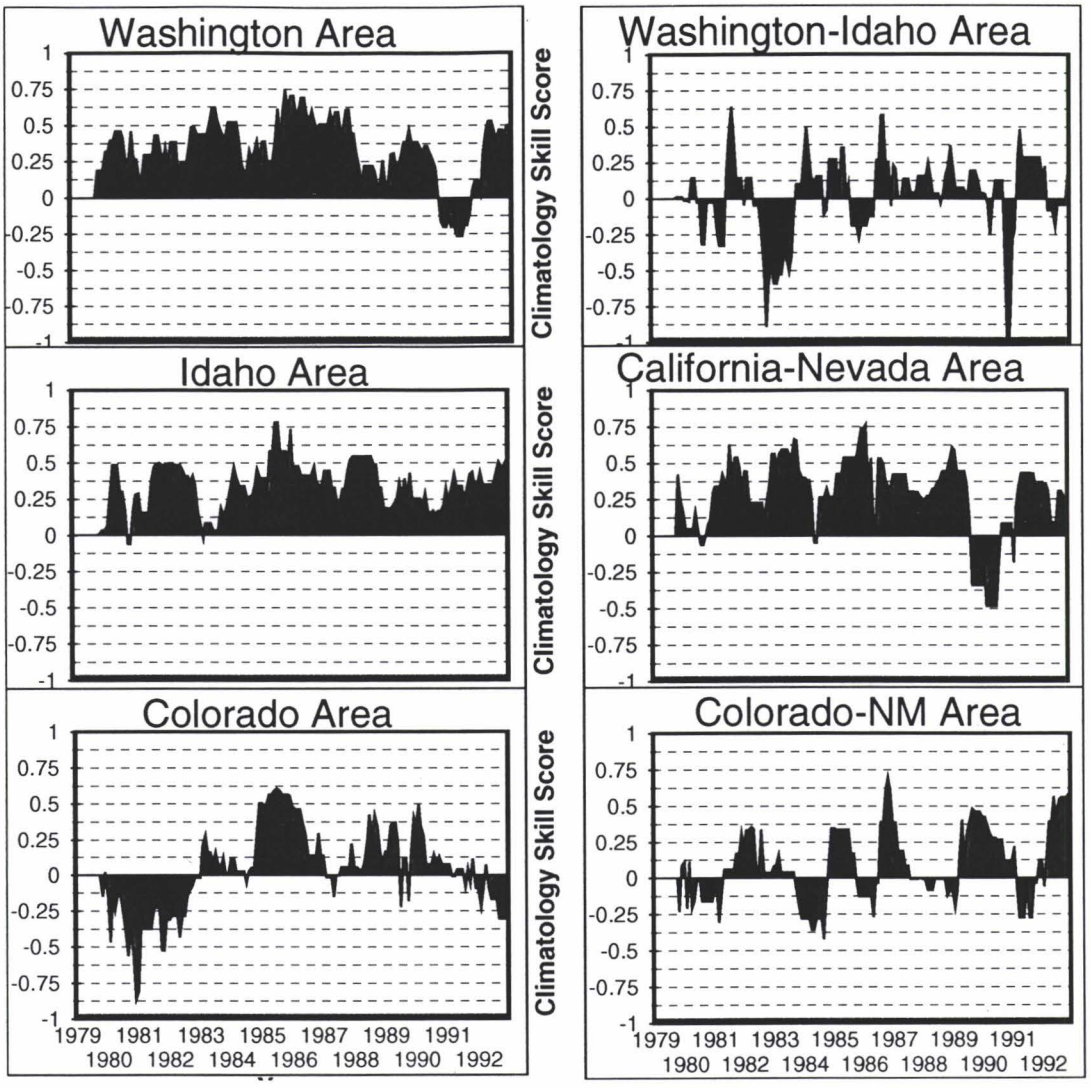

Figure 25. The model precipitation skill score. When the skill score is 1, modeled and observed data agree perfectly. At a skill score of zero, the model is predicting precipitation at the same level as the climatology. At a skill level of –1, the model is only half as good as the climatology predictor is.

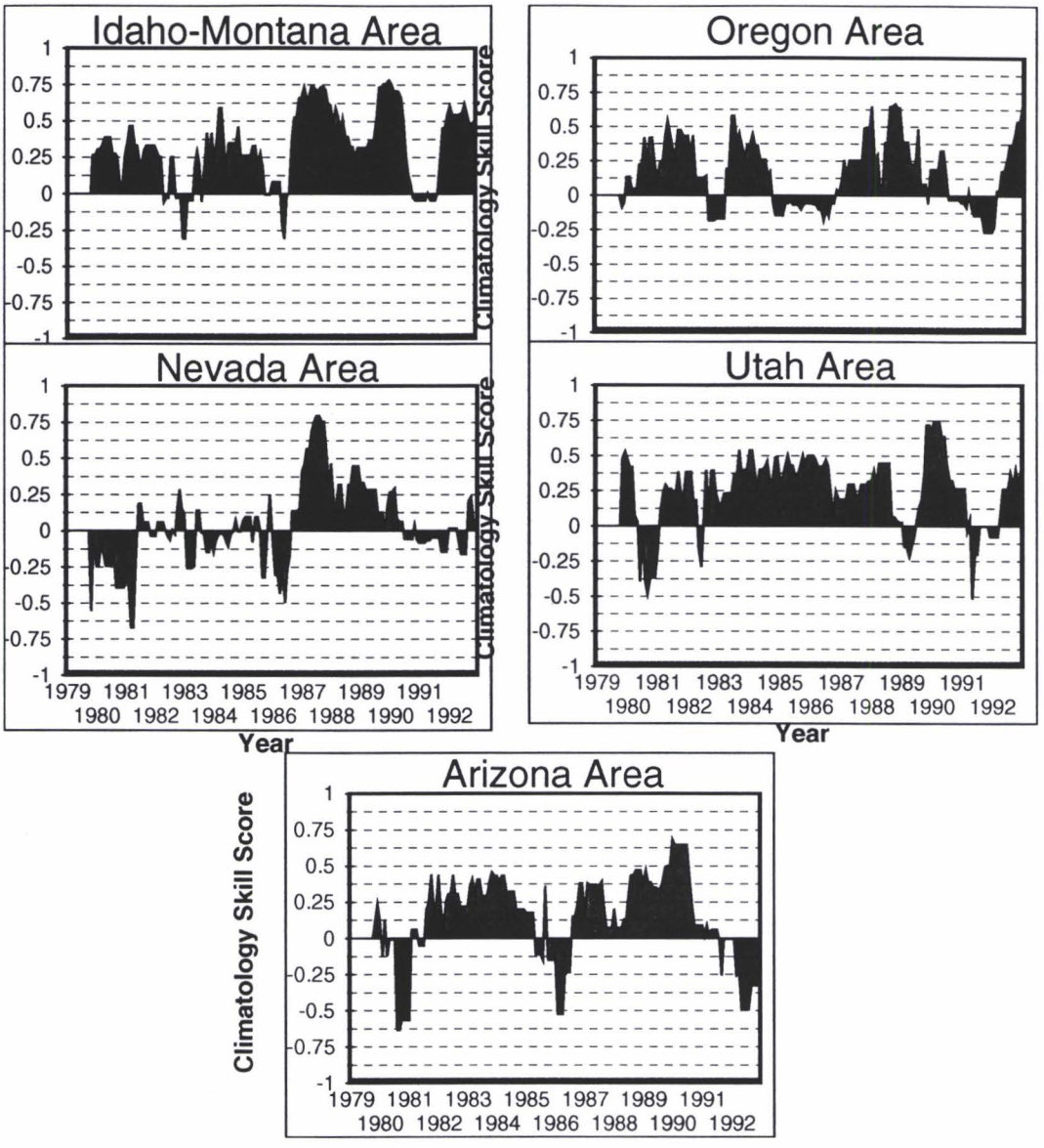

Figure 26. The model precipitation skill score. When the skill score is 1, modeled and observed data agree perfectly. At a skill score of zero, the model is predicting precipitation at the same level as the climatology. At a skill level of –1, the model is only half as good as the climatology predictor is.

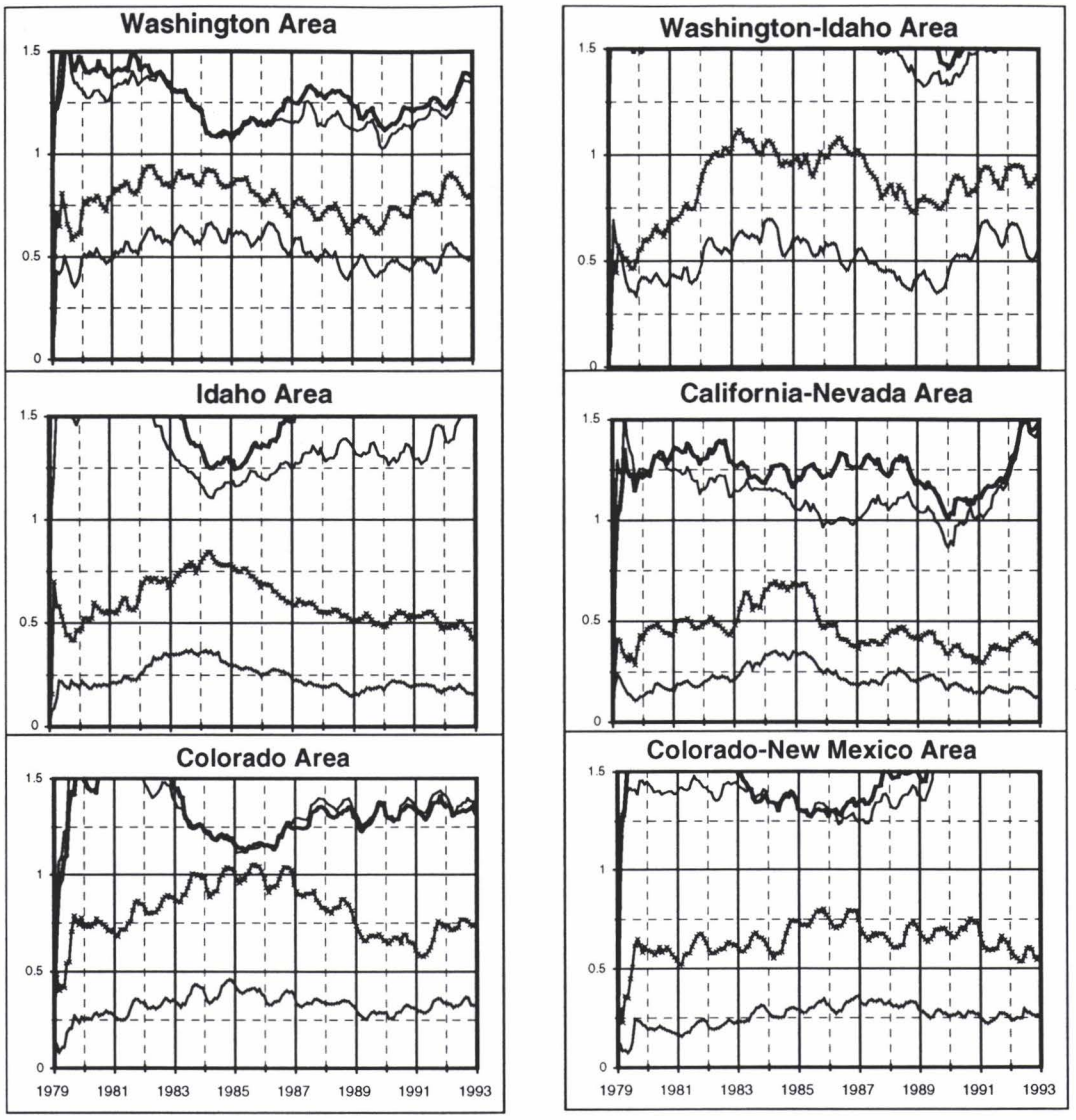

Figure 27. Precipitation bias score generated using daily data smoothed by a 30 day running mean. The bias score is the model/observed ratio of events of a given magnitude averaged over the 25 points for each day.

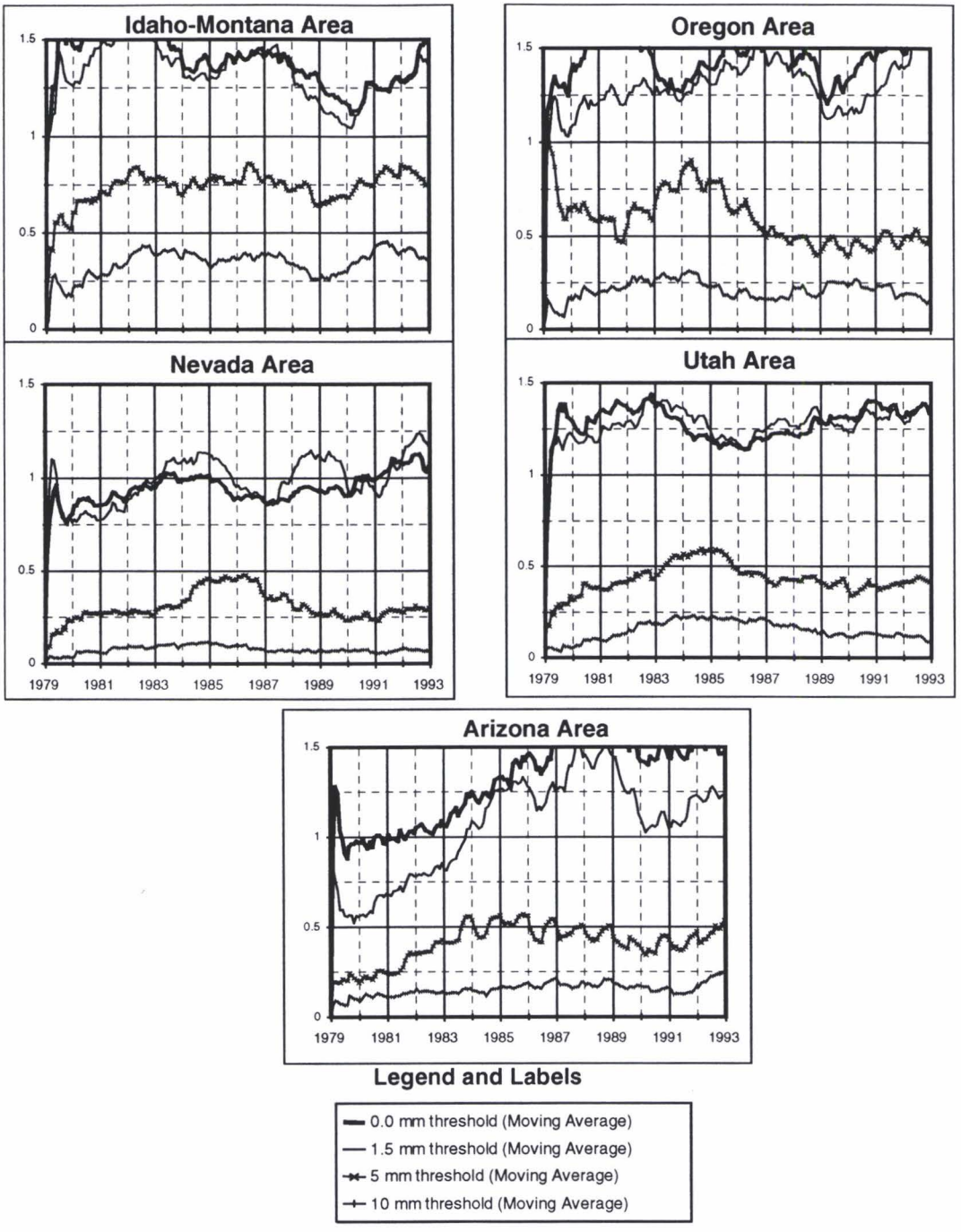

Figure 28. Precipitation bias score generated using daily data smoothed by a 30 day running mean. The bias score is the model/observed ratio of events of a given magnitude averaged over the 25 points for each day.

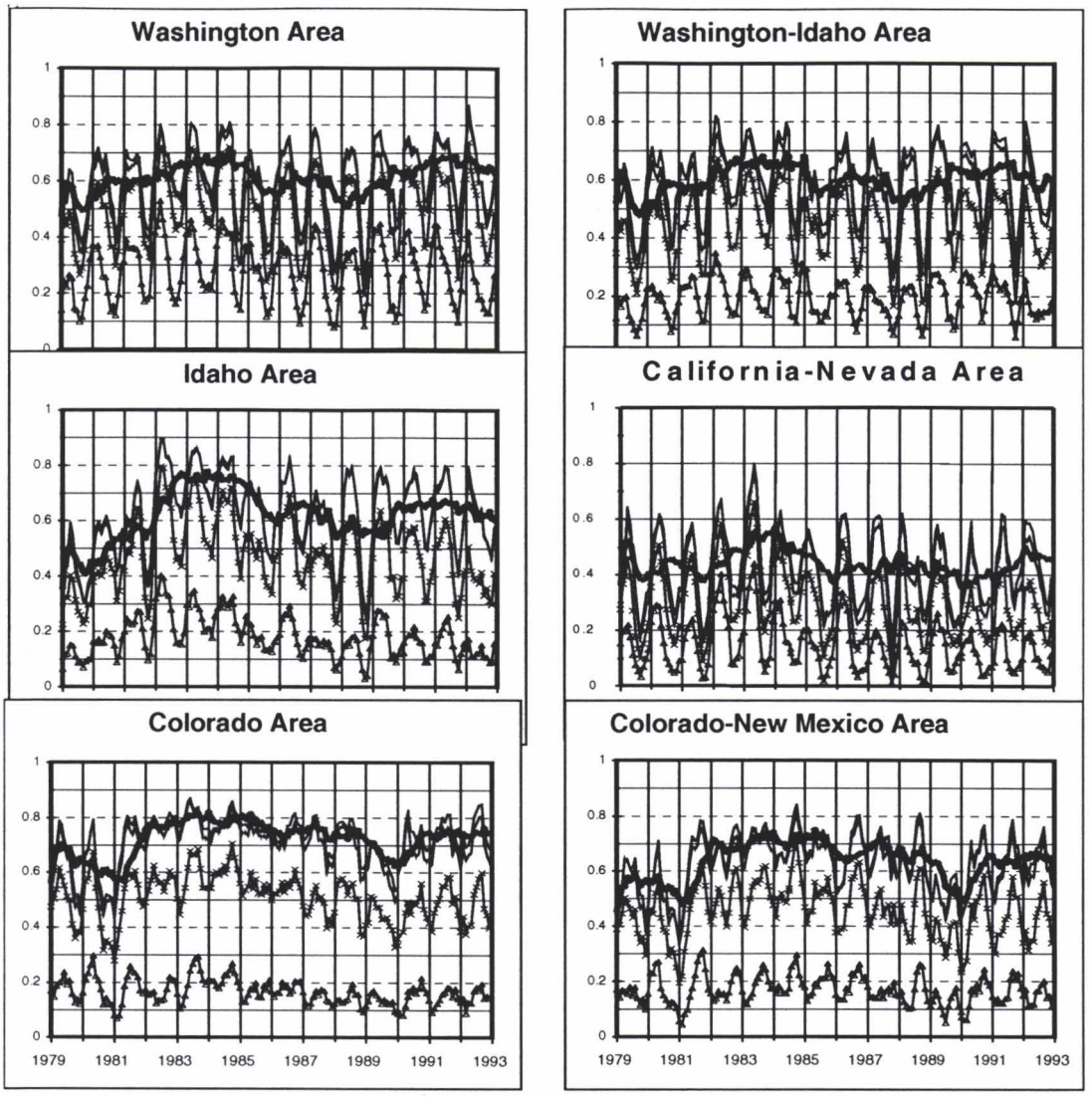

Figure 29. . Precipitation threat score analysis of the model output, compared with the observed data. When TS = 1, the model is perfect.

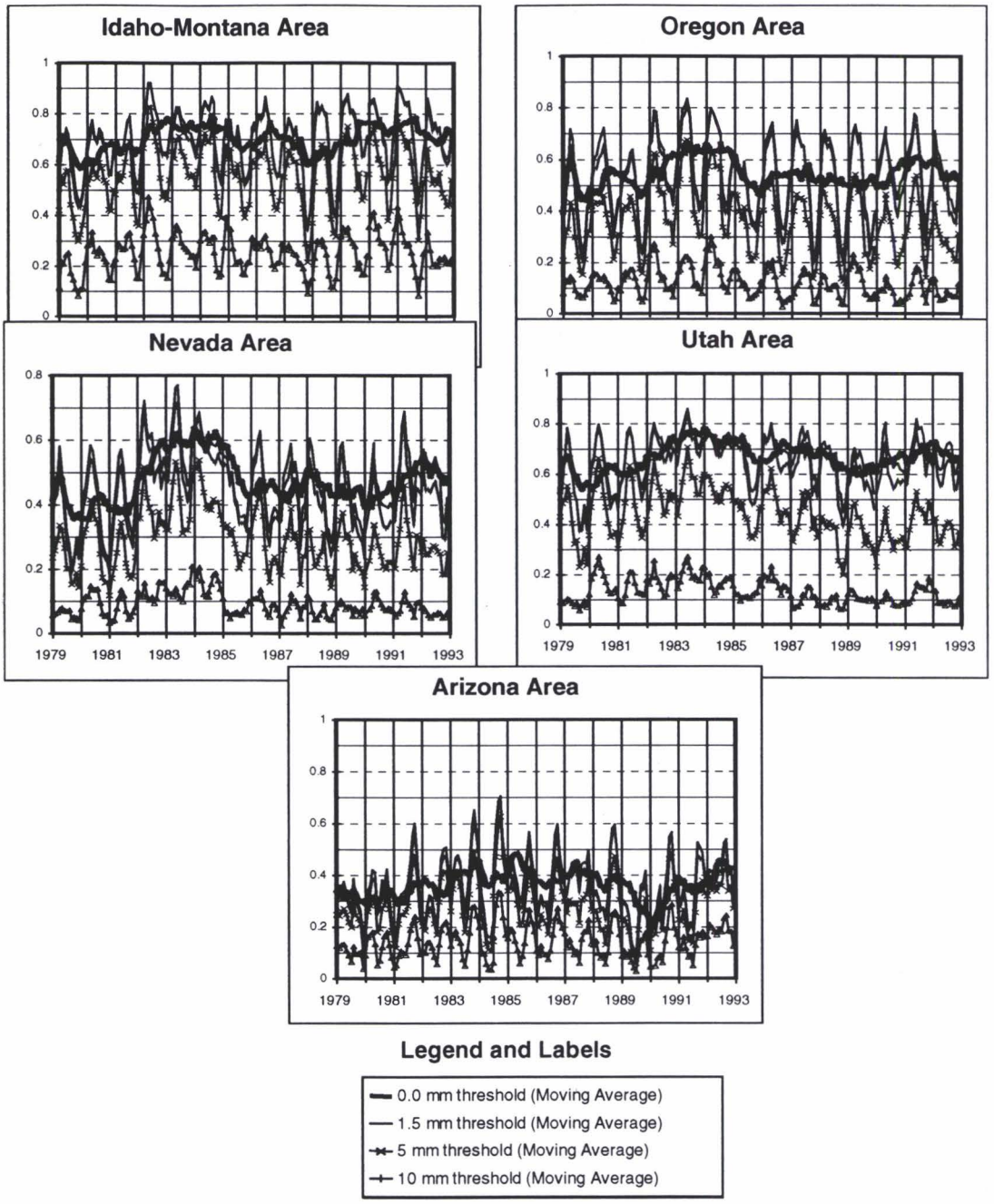

Figure 30. Precipitation threat score analysis of the model output, compared with the observed data. When TS = 1, the model is perfect.

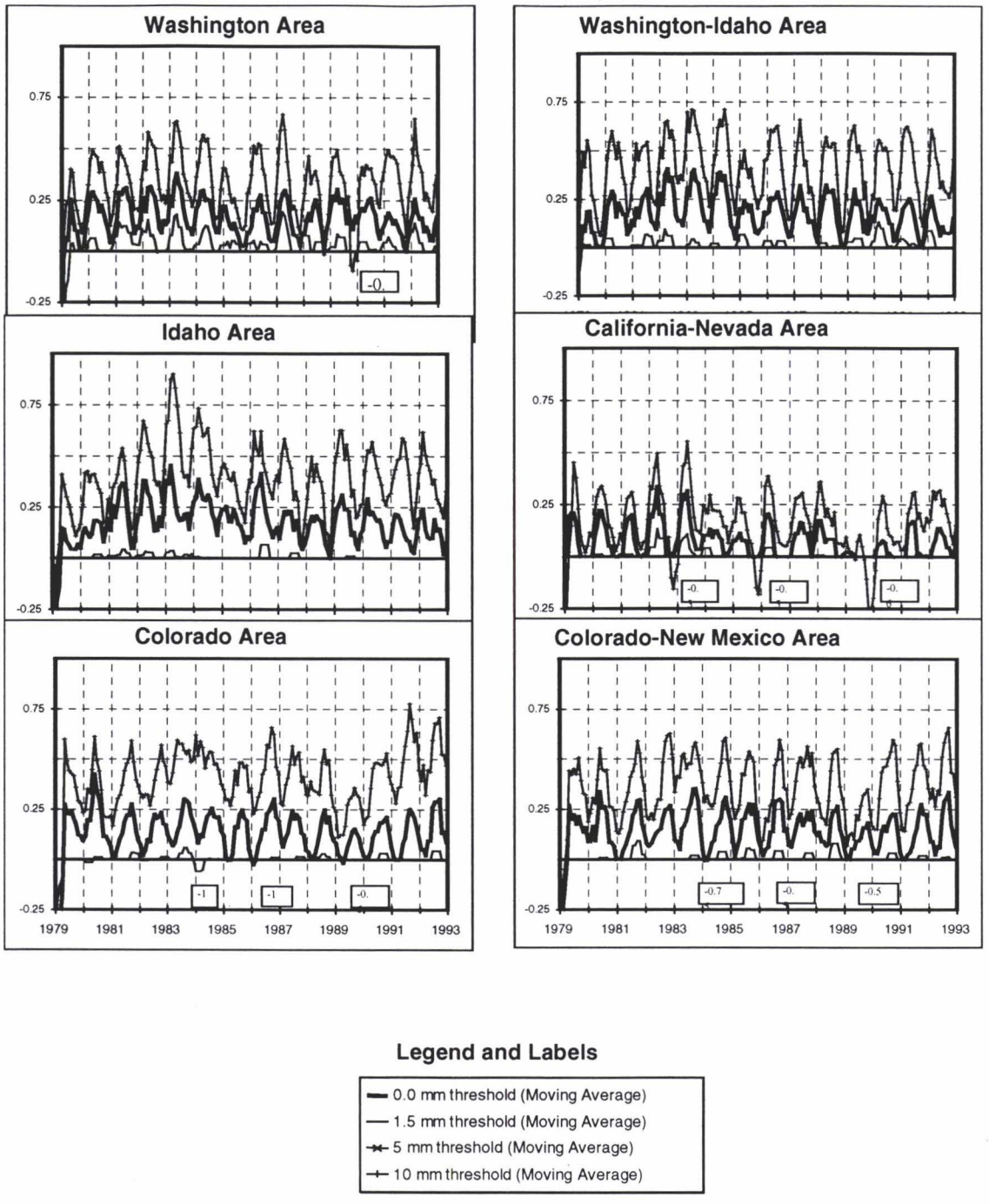

Figure 31. Precipitation threat score analysis of the model output, compared with the Monte Carlo sampling of the observed data. When $TS_{mc} > 0$, the model has predictive capability.

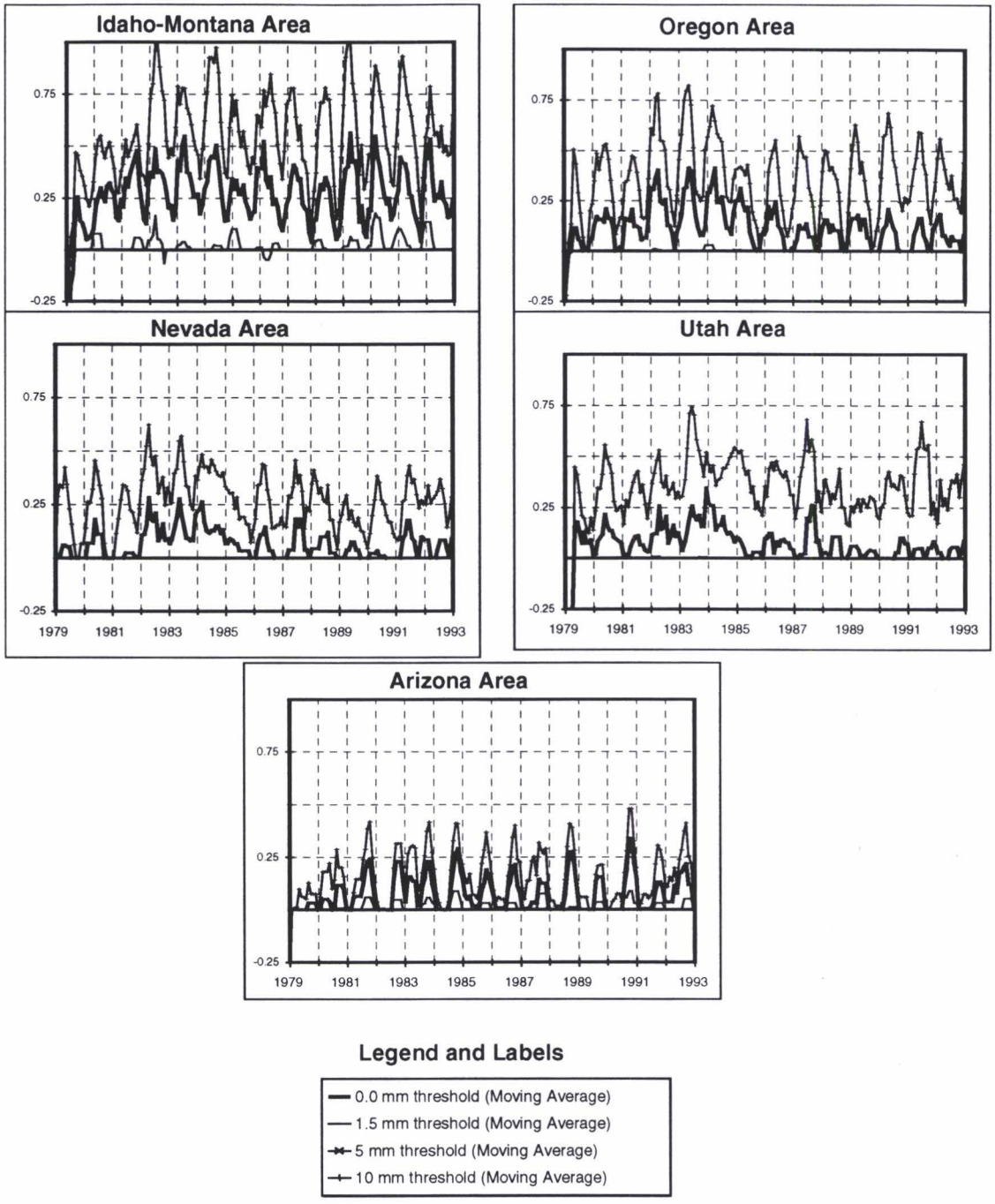

Figure 32. Precipitation threat score analysis of the model output, compared with the Monte Carlo sampling of the observed data. When $TS_{mc} > 0$, the model has predictive capability.

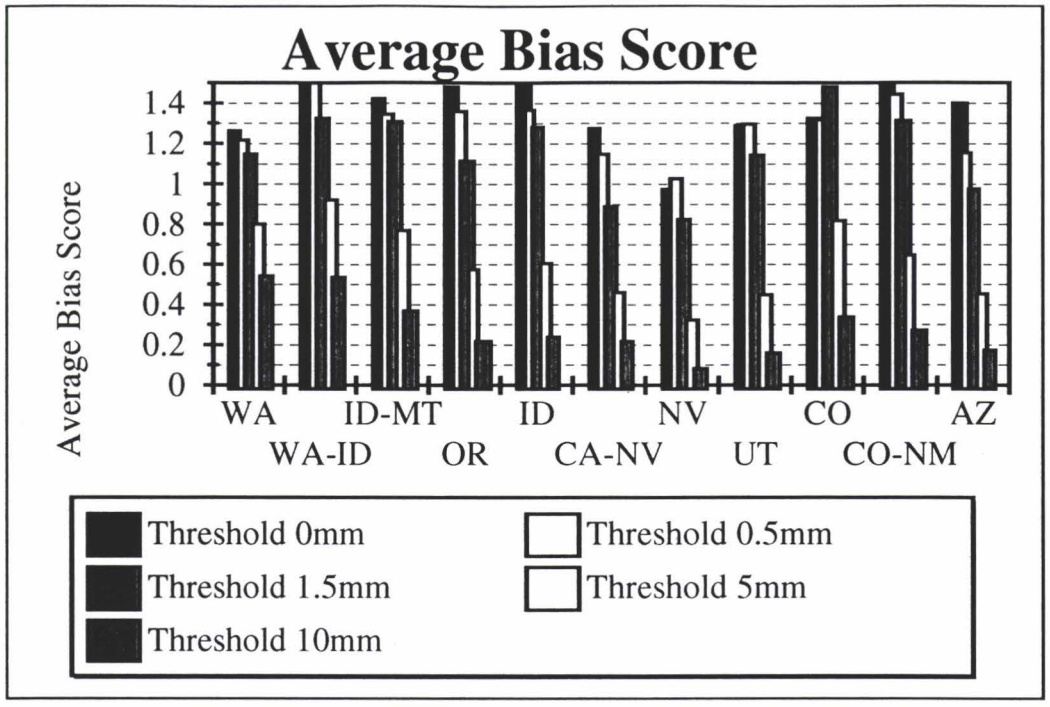

Figure 33. Annual average precipitation bias score.

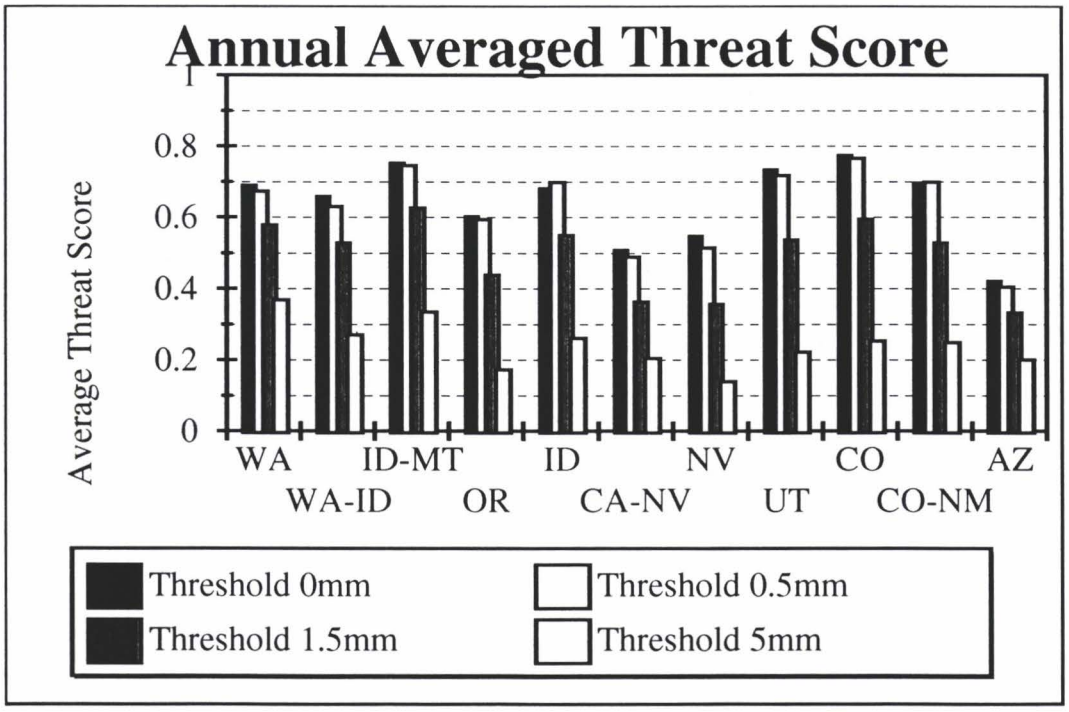

Figure 34. Annual average precipitation threat score.

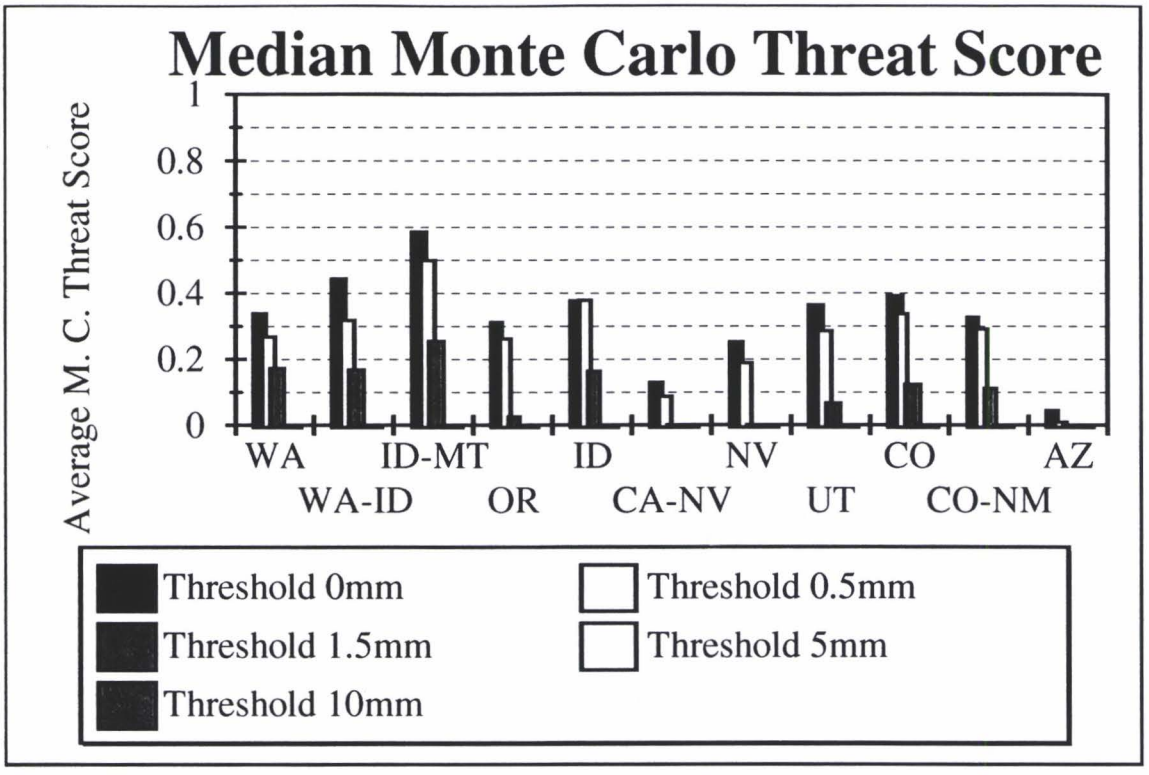

Figure 35. Annual average precipitation Monte-Carlo threat score.

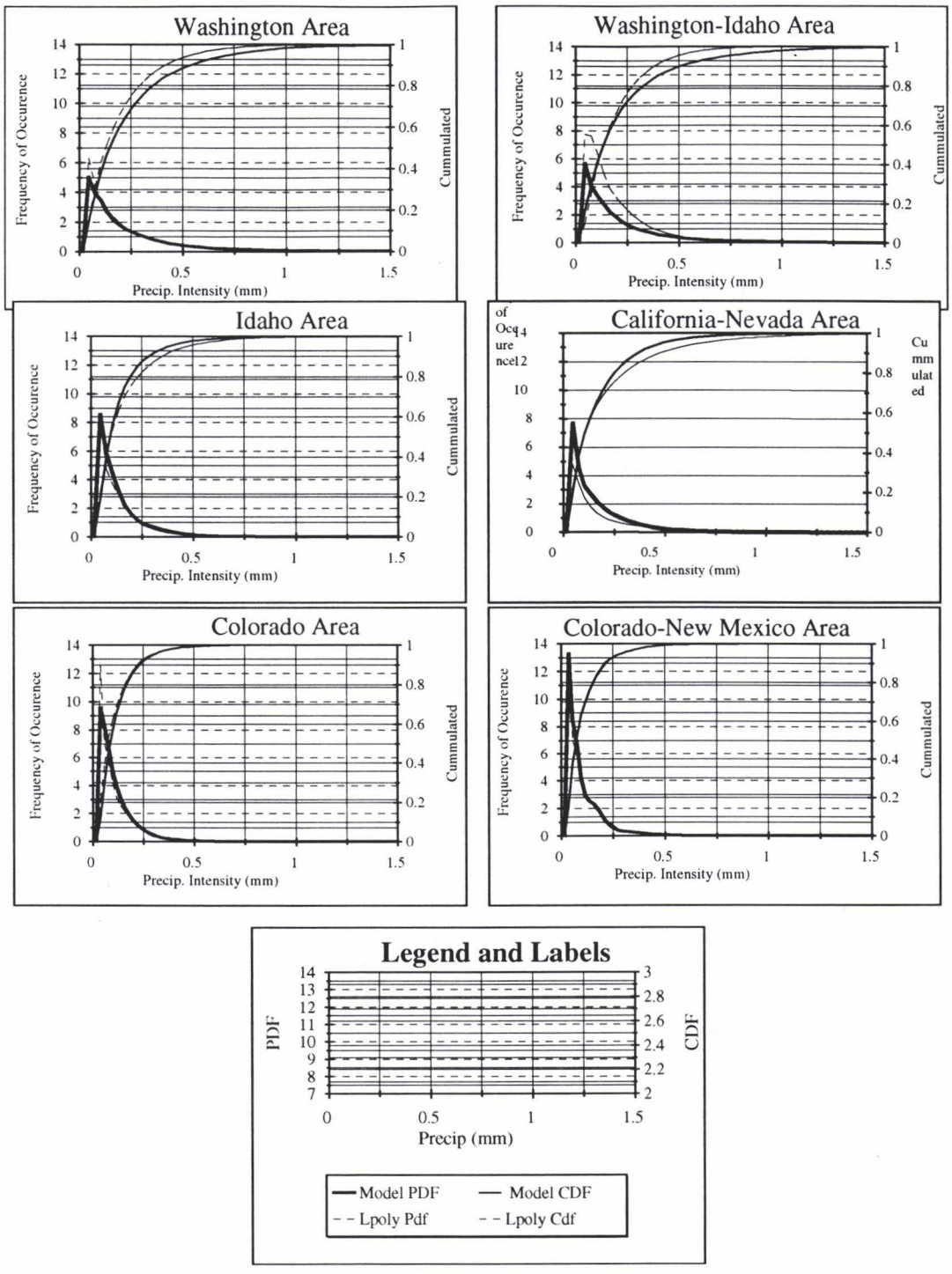

Figure 36. Winter time (December, January, and February) precipitation intensity distribution calculated for the 14-year period of study. Cumulated frequency is normalized to 1.

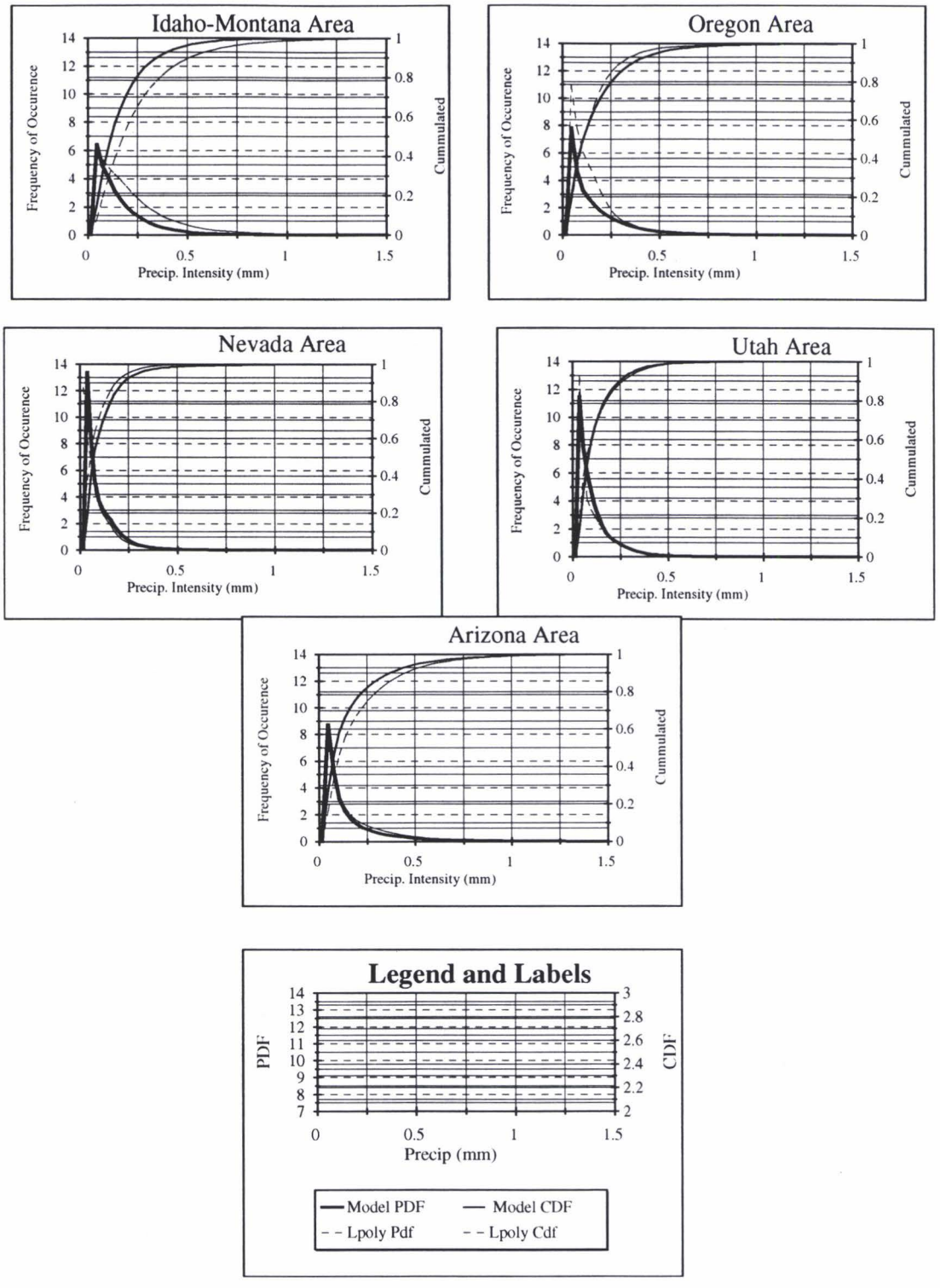

Figure 37. Winter time (December, January, and February) precipitation intensity distribution calcualted for the 14-year period of study. Cumulated frequency is normalized to 1.

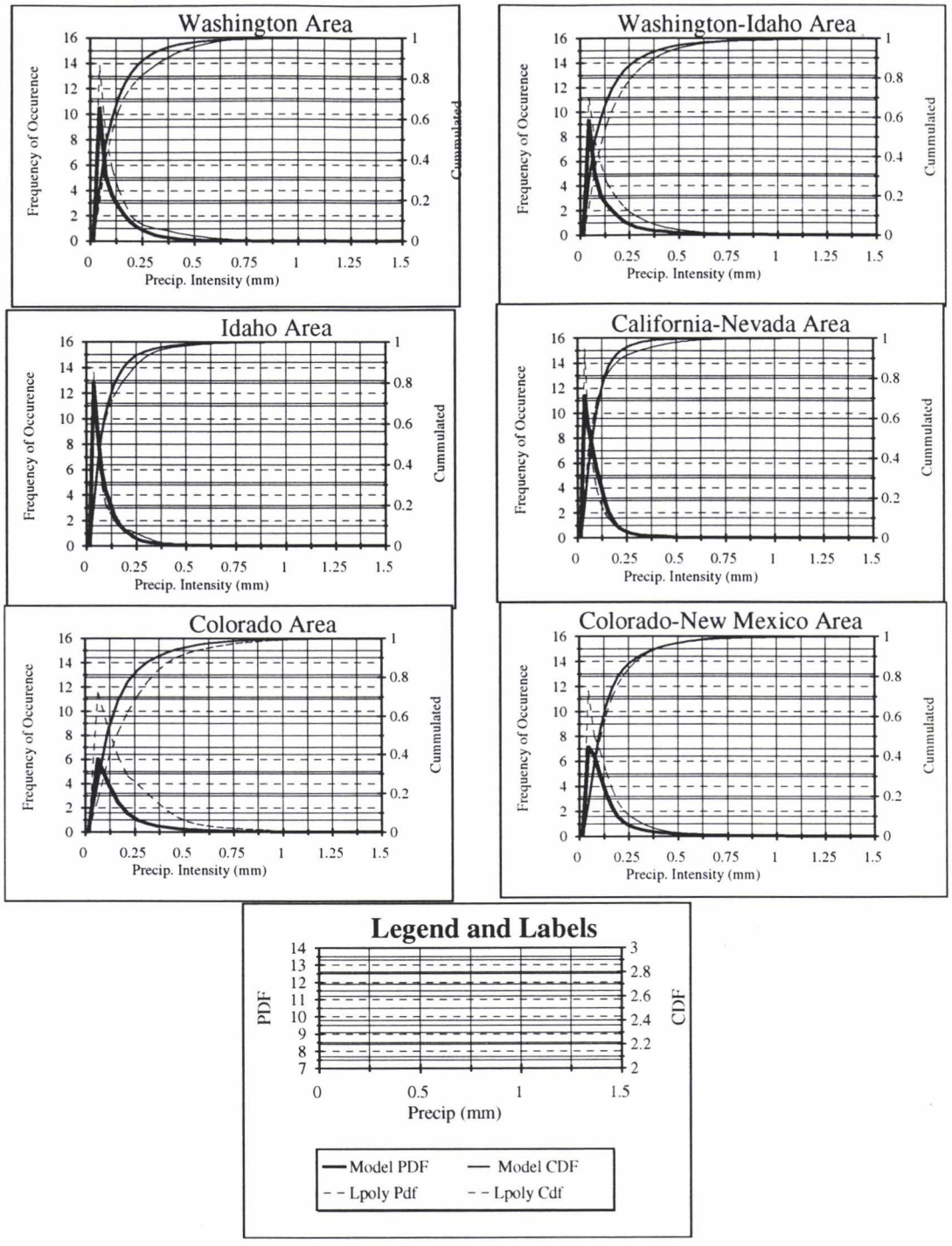

Figure 38. Summer precipitation intensity for six of the eleven analysis areas in the western US. Cumulated frequency is normalized to 1.

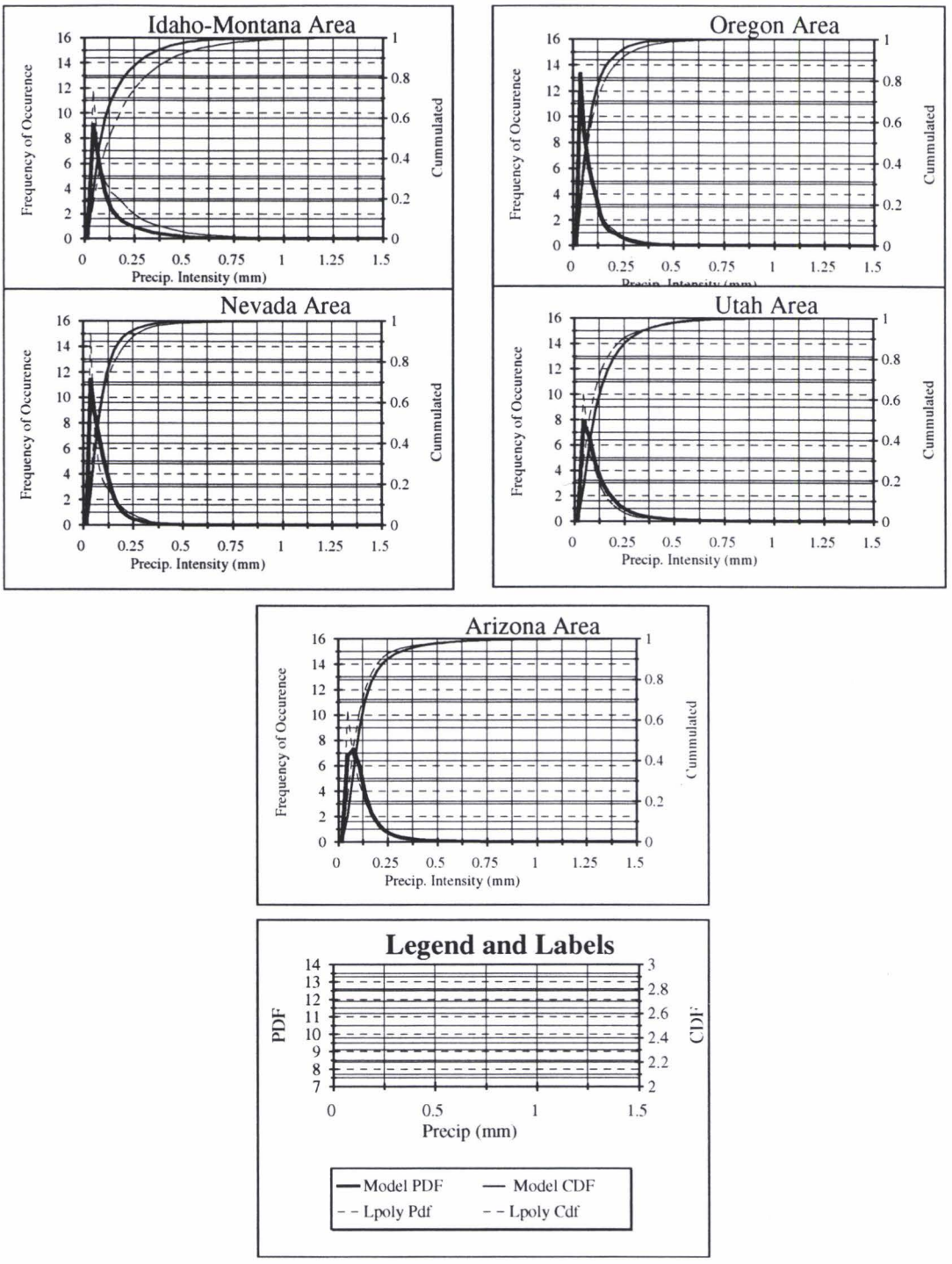

Figure 39. Summer precipitation intensity for five of the eleven analysis areas in the western US. Cumulated frequency is normalized to 1.

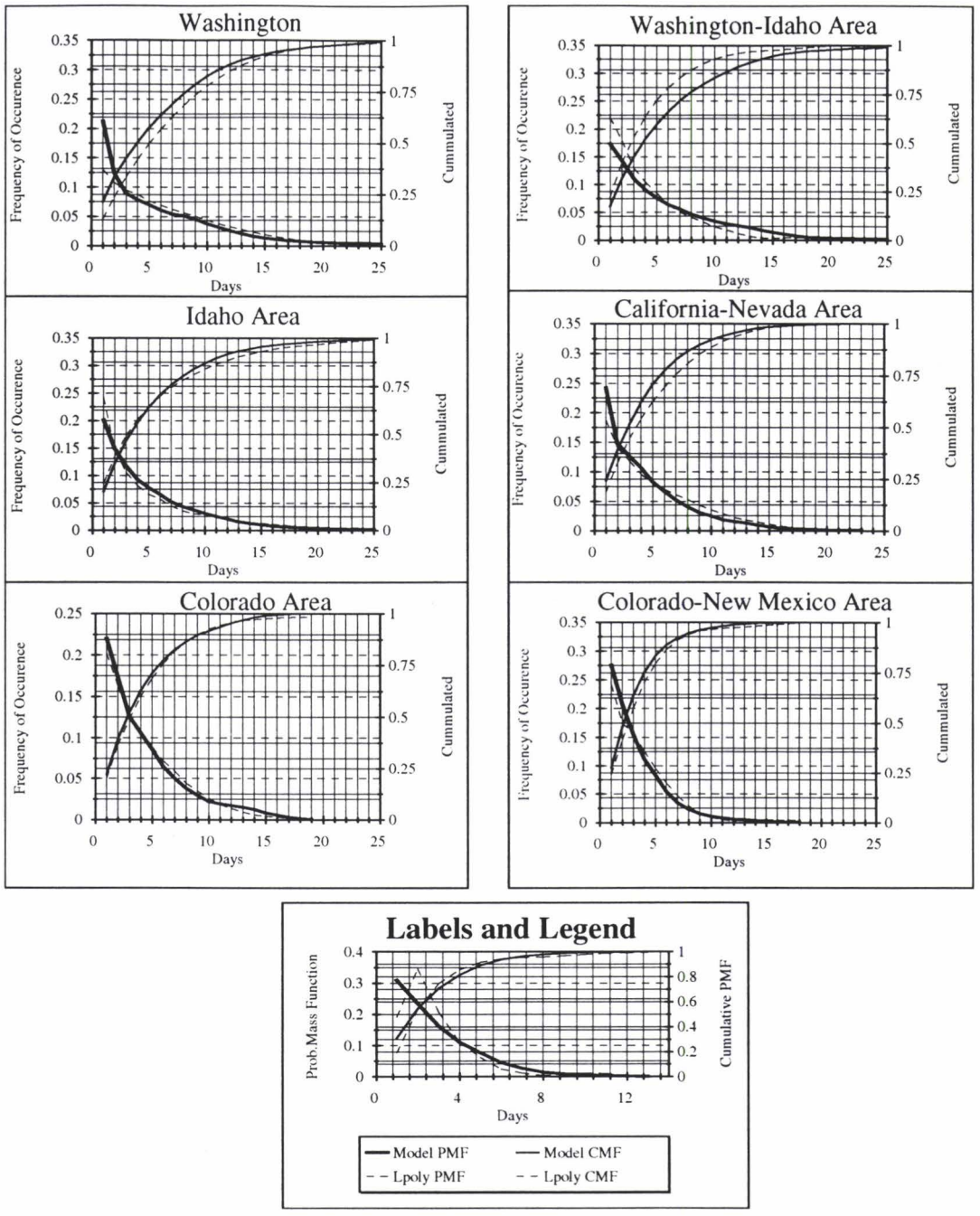

Figure 40. The wet spell length distribution comparison for the observed and model for the winter period (December, January, and February). Cummulated frequency is normalized to 1.

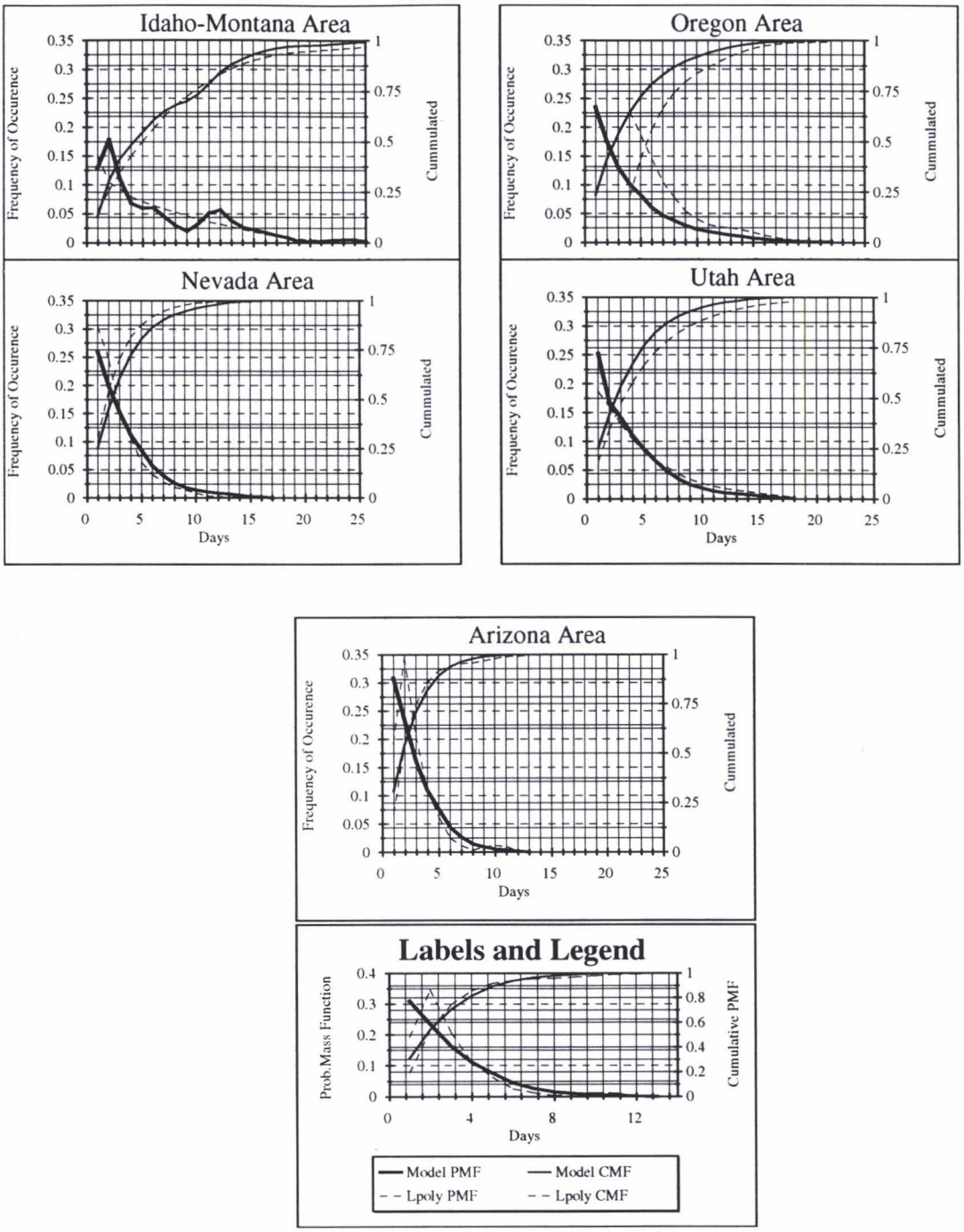

Figure 41. The wet spell length distribution comparison for the observed and model for the winter period (December, January, and February). Cummulated frequency is normalized to 1.

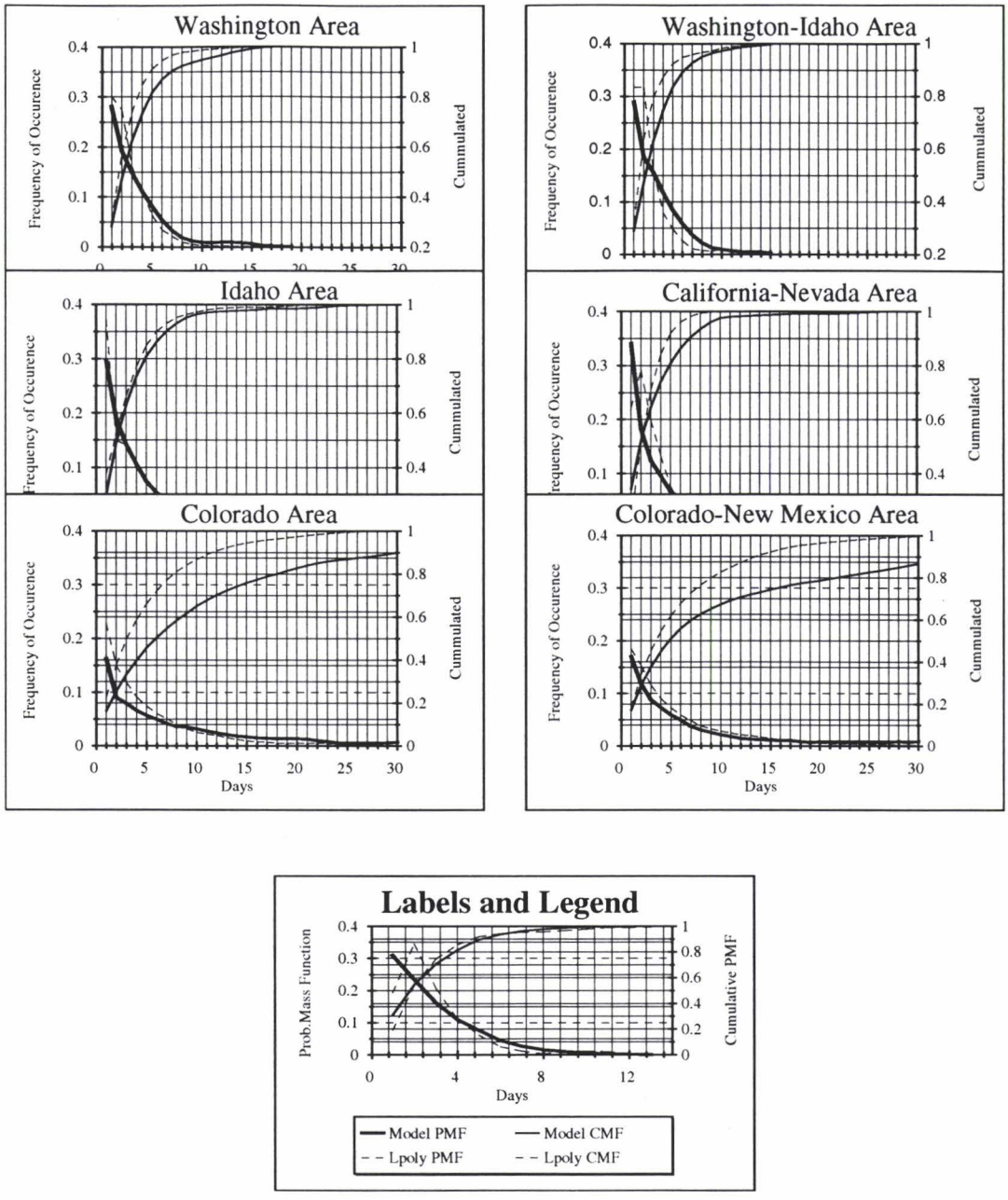

Figure 42. The wet spell length distribution comparison for the observed and model for summer period (June, July, and August). Cummulated frequency is normalized to 1.

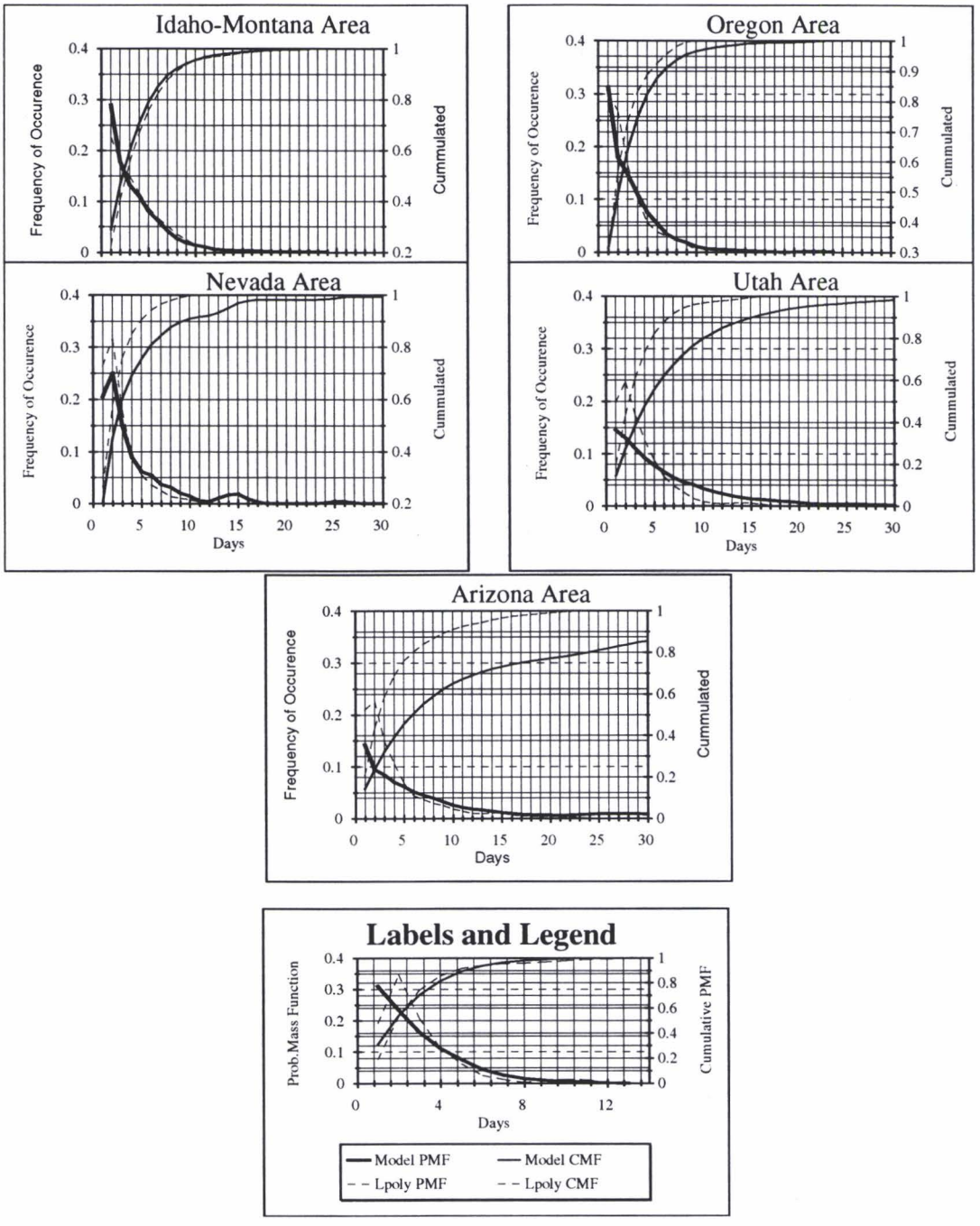

Figure 43. The wet spell length distribution comparison for the observed and model for summer period (June, July, and August). Cummulated frequency is normalized to 1..

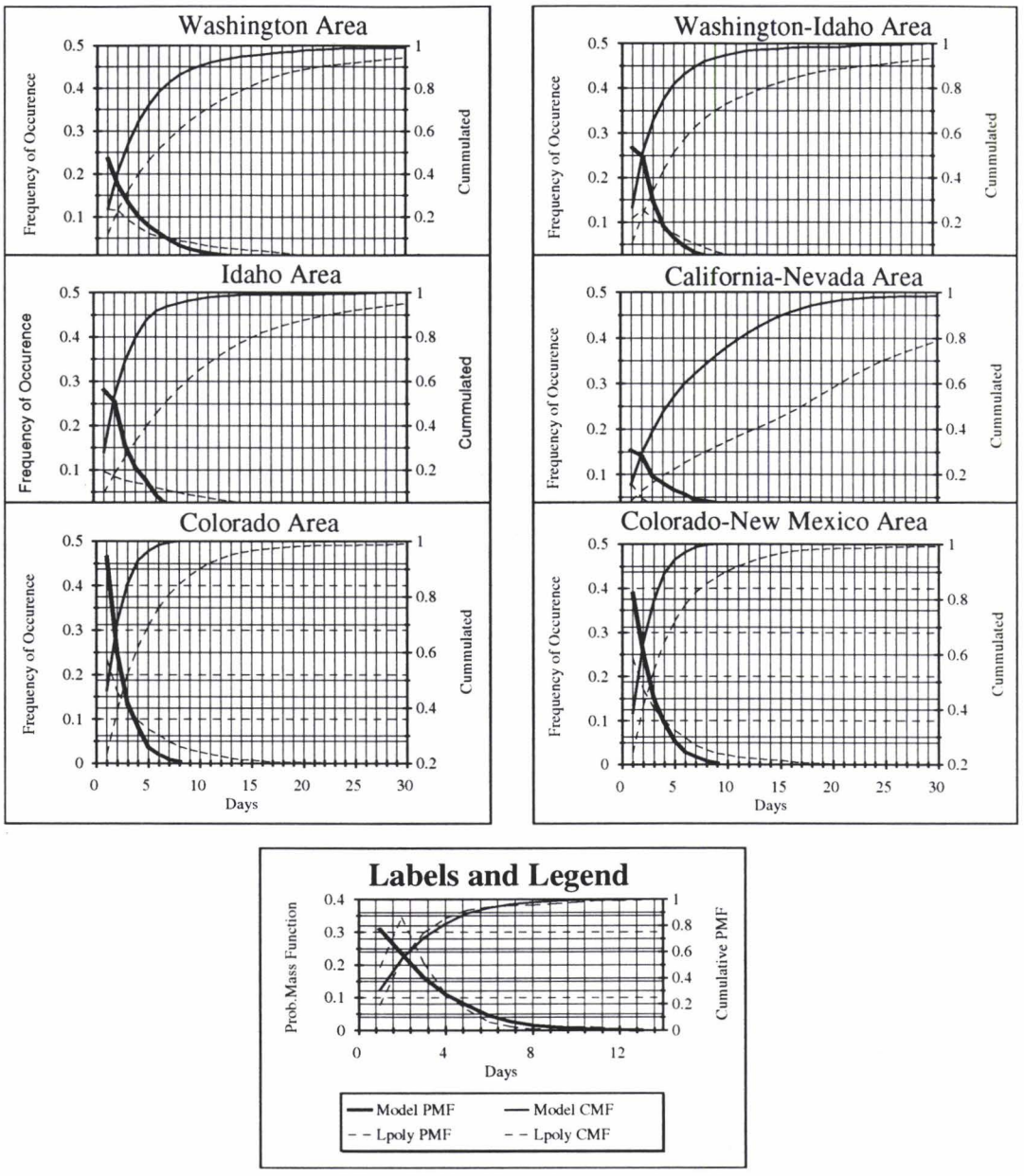

Figure 44. The dry spell length distribution comparison for the observed and model for summer period (June, July, and August). Cummulated frequency is normalized to 1.

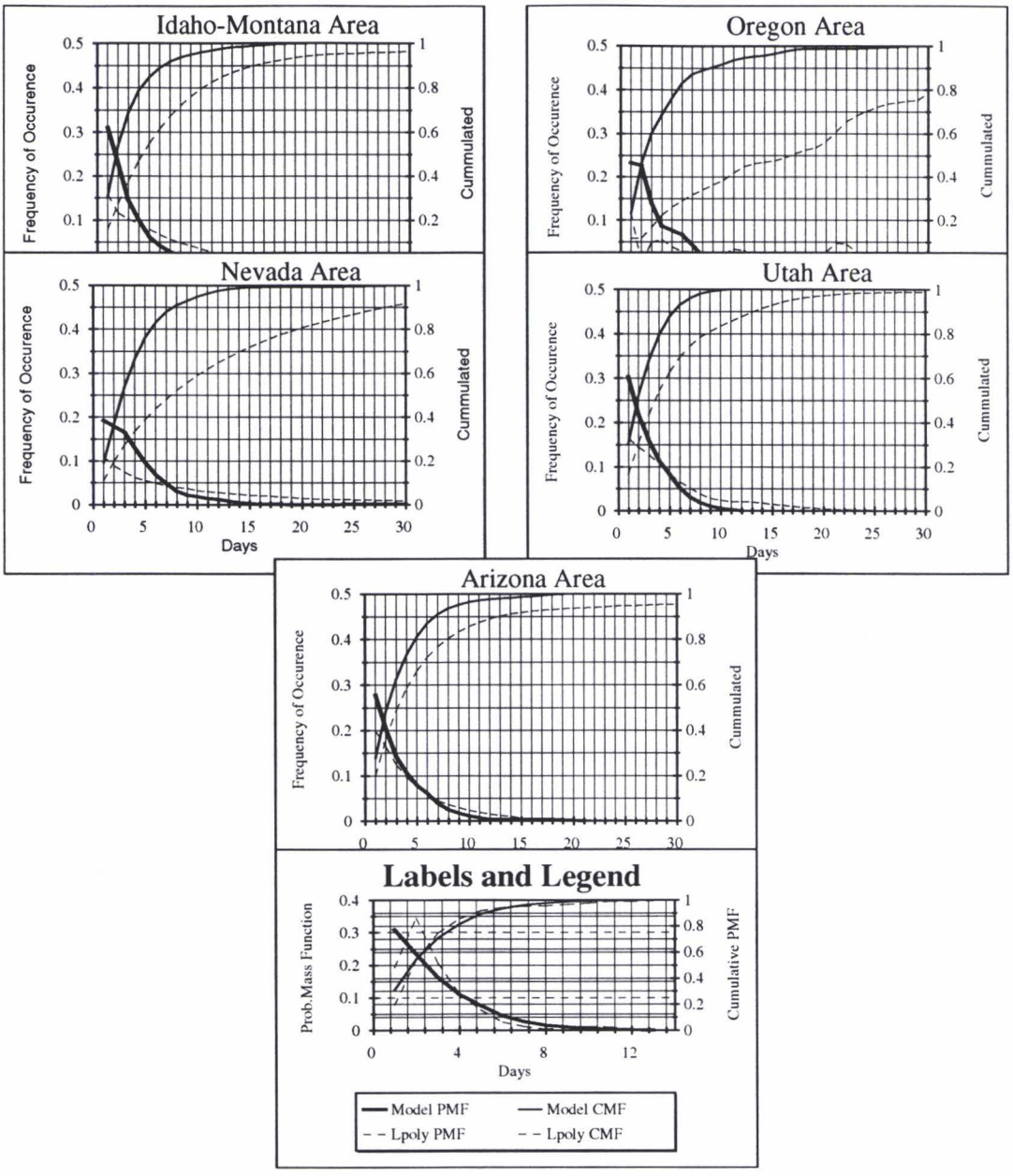

Figure 45. The dry spell length distribution comparison for the observed and model for summer period (June, July, and August). Cummulated frequency is normalized to 1.

CHAPTER IV: THE IMPACT OF LATERAL BOUNDARY CONDITIONS ON REGIONAL CLIMATE MODELING: A HIGH RESOLUTION TEST CASE IN COMPLEX TERRAIN

Abstract

Recently there is interest to integrate regional climate models at higher resolutions and for longer periods than in the past. In some cases there is considerable work required before a model can produce successful simulations and in some cases the problems are unique to the resolution. Using the regional model RegCM2 we found the domain average precipitation to be lower than observed, and precipitation too tightly bound to mountain up-slopes. Although some of these problems may be unique to RegCM2 or our domain, we believe their solutions are linked to regional modeling at 10 km resolution. We suggest that at this resolution regional models require explicit cloud and rain water microphysics with some modifications. For small domains, cloud and rain water should be input on the lateral boundary conditions and update frequency may need to be increased. We present and discuss extensive analysis of our simulations.

IV.1 INTRODUCTION

There is continued interest in high-resolution climate analysis for various impact studies including hydrological studies of local floods and droughts, and studies of vegetation, forest, and crops. With increased concerns of long-term large-scale weather patterns such as El-Nino Southern Oscillation (ENSO) and climate change, there is even more concern with the regional impact from large-scale climate variation. Researchers and decision-makers want to resolve not only what happens on a global scale, but how large-scale changes affect a particular country, state, political district, or watershed. There are also concerns with atmospheric chemistry and pollutants, the impact of an increasing population, and the feedback a changing climate may have. With new legislation on the global release of greenhouse gases, there will continue to be interest in determining how these changes affect the local climate. There is also a concern that many of the stresses that limit plant growth disappear in global climate model (GCM) studies because of climatic averaging [*Paoli,* 1994]. With computing speeds continuing to increase, GCMs are being run at higher and higher resolutions. Likewise, limited area model (LAM) studies must also continue to be run at higher and higher resolutions. *Kato* [1997] points out that numerical weather prediction (NWP) models are likely to be run operationally as fine as 10 km in the near future. He also points out that prior to this event the modeling community needs to "determine the optimal dynamical framework and precipitating scheme." In the case of LAMs, "dynamical framework" refers to the method of including LBCs [*Waldron et al.,* 1996]: spectral methods, continuously varying grid-size, Davies nudging [*Davies and Turner*, 1977], LBC from outer model, plus large-scale forcing internally from the outer model, or two-way nesting. Precipitating scheme refers to the method of solving for precipitation. There are many choices for this (implicit vs. explicit cloud and rain water, as well as sub-grid-scale convective parameterization). The same need to determine the "optimal dynamical framework and precipitating scheme" also applies to regional climate models. The mechanics of nesting LAM beneath higher resolution GCM runs need to be worked out. A particular challenge for determining the precipitation scheme is that the solution depends on resolution [*Molinari and Dudek*, 1992]. *Molinari and Dudek* [1992] point out that

there is not an obvious choice for the precipitation scheme at 10 km resolution. This underlines the need for more experimentation at this level. Regional climate models are good tools for bringing fine-scale information from a GCM simulation. The modeling community is now at the point where there is an interest to run a LAM on the interannual timescale at finer resolution than has been tried before. This study examines some of the scientific problems encountered when a LAM, traditionally run at resolutions on the order of 50 km, is used at a horizontal resolution of 10 km. The most significant error, and perhaps the most important, is the overprediction of precipitation on mountain up-slopes, coupled with an underprediction of precipitation elsewhere [*Nickerson et al.*, 1986; *Zhang et al.*, 1988]. We present analysis of the problem for our simulations and propose possible sources for a solution. We propose that there exist certain deficiencies in the RegCM2 explicit cloud microphysics. We also propose that cloud and rain water need to be input on the LBCs for small domains. Finally, we present some problems that occur when LBCs are not updated at high enough of a frequency. More precisely, we use the National Center for Atmospheric Research's (NCAR) LAM climate model Regional Climate Model version 2 (RegCM2) [*Giorgi et al.*, 1993a, 1993b]. We do this at 10 km resolution over the complex terrain of the northern Utah Rocky Mountains. See Figure 8 for a plot of our 50 km model grid, and Figure 46 for the nested 10 km model grid. The 10 km grid includes the Great Salt Lake, the Uinta, Wasatch, and Bear River mountain ranges, and parts of Idaho and Wyoming. Also included is the Salt Lake City rawinsonde that is near the center of the domain. The results shown are for only one month, September 1981. Further simulations will need to be performed to verify the conclusions of this study for other areas and seasons. We provide the theoretical underpinning for our results based on basic physics and we believe the results can have validity outside the area and season demonstrated.

IV.2 DISCUSSION

The RegCM2 [*Giorgi et al.*, 1993a, 1993b] model is a hydrostatic, compressible, sigma layer, primitive equation atmospheric model. The basic model is mesoscale model version 4 (MM4) as described by *Anthes et al.* [1987]. It has sophisticated packages for surface physics (Biosphere Atmosphere Transfer Scheme version 1e (BATS1E)), [*Dickinson et al.*, 1993], lakes [*Hostetler and Giorgi*, 1992], and a non-local atmospheric boundary layer [*Holtslag et al.*, 1990; *Holtslag and Boville*, 1992]. A delta-Eddington approach to solar radiation [*Briegleb*, 1992] and an option to use the Grell convective parameterization scheme [*Grell et al.*, 1994] – which we used with a *Fritsch and Chappell* [1980] type closure – were included. RegCM2 is an improved version of the RegCM model used in several earlier studies [*Giorgi and Bates*, 1989; *Giorgi*, 1990, 1991; *Medina*, 1991; *Giorgi et al.*, 1992; *Matthews et al.*, 1992].

To save CPU time, RegCM2 replaced the Brown-Campana time differencing approach used in MM4 [*Anthes et al.*, 1987] with a split-explicit approach [*Madala*, 1981]. In this study, two model domains are used – a 50 km model domain and a nested 10 km domain. The first model domain (Figure 8) is similar to the area over the western US studied by *Giorgi et al.* [1992], except that a grid domain of 72 x 60, 50 km square boxes, with 17 vertical layers is used. We apply terrain smoothing on the lateral boundaries, similar to *Hong and Juang* [1998]. The lowest model layer is approximately 20 m off the surface and the top model layer is 80 hPa. In the second model domain (Figure 46) the grid size is 36 x 37, 10 km grid-squares with 20 vertical layers. Like the 50 km simulation, we apply terrain smoothing on the lateral boundaries. For these simulations we use the explicit cloud and rain water scheme contained in RegCM2 [*Hsie and Anthes*, 1984].

In Figure 47 and Figure 48, we show the gridded observed precipitation and error over the 10 km grid domain for September 1981. The observed precipitation is gridded using a local polynomial regression in latitude, longitude, and elevation [*Lall and Bosworth,* 1994]. This is the same methodology as used in chapter II for the 50 km data. In this case monthly station data are used to generate the dataset, rather than gridding daily data and accumulating the monthly total. Since monthly precipitation is much more smoothly distributed over the domain, it is easier to grid and the errors are smaller than daily precipitation (see chapter II). However, in this case the number of grid-points (1260) is an order of magnitude greater than the number of station observations (140), and we must use a little caution. Figure 49 shows the location of the station observations and the grid-points. As shown, most grid-points do not have data within them or even in grid-points nearby. As such, the error estimates are much too low since they do not take this into account. Further experimentation with the LP found that a small change in the minimum number of points used in the regression allowed a specific grid-point in the Uinta Mountains to reach unrealistic values (120 mm). This points out the difficulty with extrapolating a high density gridded dataset from a low density observation network. However, as this is the best that can be done, we will use this dataset for comparision to the model experiments.

Figure 50 shows the RegCM2 modeled precipitation for the same period as the observed precipitation shown in Figure 47. Notice that the precipitation is too highly concentrated on the mountain up-slopes and much too low over the entire grid domain. Also the simulation has five epicenters of tightly compact precipitation, and the observed is much more diffuse. However, the fact that the observed is diffuse probably has more to do with the limited amount of observed data stations in the region. This behavior is similar to the 10 km simulation of *Nickerson et al.* [1986], and the 12.5 km simulation of *Zhang et al.* [1988]. The Nickerson simulation was only a one-day simulation and they theorized that problems with the rainfall distribution had to do with spin-up. Zhang theorized that a 12.5 km simulation was too high of a resolution for the hydrostatic model, and wondered if the vertical velocities were too low [*Zhang et al.,* 1988]. Our simulation was long enough to eliminate concerns with spin-up and therefore we theorize that problems with rainfall distribution are not related to this. Simulations over the complex terrain of Japan by *Kato* [1997] showed that while a 10 km hydrostatic simulation exaggerated maximum upward velocities, the impact on precipitation was very small. The net effect was an increase in precipitation amounts by 8%, and area by 12%. Thus we conclude that the use of a non-hydrostatic model, compared to the use of our hydrostatic model, would not impact results significantly.

Let us investigate the possible causes of this increased distribution of precipitation over mountain up-slopes. As pointed out by *Sass and Christensen* [1995], the quality of a LAM simulation is dependent upon three factors – the quality of the LAM itself, the quality of the boundary relaxation scheme, and the quality of the boundary fields. We will investigate four issues: issues regarding the explicit cloud microphysics and cumulus parameterization, the need for cloud and rain water to be included on the LBCs, the quality of the input LBCs, and finally the LBC update frequency. To establish these points, we first examine theoretical issues associated with resolution and modeling of convective systems. Next we assess the quality of the LBC data input into our model simulations. Lastly we examine some theoretical implications of the update frequency for LBC applied on a LAM. Finally, we will verify the conclusions of the theoretical analysis with the results of our model simulations.

IV.2.1 Issues with High Resolution and Modeling of Cumulus Systems

As pointed out by *Molinari and Dudek* [1992] there are difficulties associated with mesoscale modeling at resolutions between 3 km and 25 km. Convective systems, though small in scale, have an important impact on the large-scale heat and moisture budget. At these resolutions there is not the clear separation of scale for cumulus parameterization assumed in its derivation. As reiterated by *Molinari and Dudek* [1992], *Cotton and Anthes* [1989] note that sub-grid-scale convective parameterizations become "muddy" and "not well posed" below 50 km resolution. Closure assumptions are also problematic since there is no observational evidence at these very fine resolutions. Even the observational evidence at GCM scales is limited, but at the 10 km scale there is not any observational support at all. Another problem they bring out is how to simulate mesoscale organized convection, mesoscale convective complexes (MCC), where individual cumulonibus clouds organize to develop a stratiform storm system. For large-scales the MCC will remain sub-grid-scale, but for high resolutions a MCC represents a transition from the cumulus parameterization to grid-scale resolvable processes.

The next problem we will examine relates to the vertical velocities. The 10 km simulation will resolve some motions convectively parameterized at 50 km resolution. The vertical velocities will be higher and precipitation that is convectively parameterized at the lower resolution becomes grid-scale resolvable precipitation. However, since 10 km is still a large grid size, it is likely that the vertical velocities will not be high enough to produce the same amount of precipitation convectively parameterized at the lower resolution. Nor are the vertical velocities likely to be similar to that observed in cumulonumbis clouds. But, the model may produce enough grid-scale resolvable motion so as to stabilize the column [*Zhang et al.*, 1988]. Zhang recommends explicit cloud microphysics and the *Fritsch and Chappell* [1980] cumulus closure. In our initial experimentation with the model we found that when the model was run with the implicit precipitation option – where any super-saturation immediately drops out as precipitation – and cumulus parameterization, the precipitation amounts were much too high to be realistic. By using explicit microphysics, we have additional loss terms, clouds form, and cloud and rain water evaporates. Experimentation also showed that the model seemed to be insensitive to the type of closure chosen for the cumulus parameterization. Although we do not examine the fraction of precipitation that comes from the cumulus parameterization, we can only speculate that it is similar to our current simulations, where it is around 10%. We also speculate that the reason precipitation was overpredicted in the implicit case was the lack of loss terms in the explicit formulation. And, we also speculate that changes to the explicit cloud microphysics will likely allow reasonable precipitation rates for the 10 km simulation.

IV.2.2 Quality of the Input LBC Data to the 10 km RegCM2 Simulations

To establish the quality of our input LBC data, we compare these data to three datasets: a gridded surface temperature and precipitation dataset, a rawinsonde profile, and the European Centre for Medium range Weather Forcasting (ECMWF) analysis. We compare the ECMWF analysis (interpolated to the 10 km grid) to the surface dataset and rawinsonde. This validates the use of the ECMWF dataset as a verification dataset to use on the simulation input data.

IV.2.3 ECMWF Dataset Compared to Observed

The ECMWF dataset is first interpolated to the 10 km grid domain. Next we find the grid-point in the 10 km grid domain nearest to the Salt Lake City rawinsonde, and compare the two. The ECMWF point nearest to the Salt Lake City (SLC) rawinsonde is within 2 degrees Celsius in temperature, 3 m/sec in winds, 3.0 g/kg in specific humidity, and within 20% in relative humidity. See Figures 51, 52, 53, and 54 for a comparison of the ECMWF data-point to the SLC rawinsonde. Most of the differences are within the variability of the observed dataset. However, the ECMWF is significantly too moist near the surface, and the variabilities in temperature and wind speed are too low near the surface. The variability in relative humidity is also too low.

IV.2.4 Simulation Input Dataset Compared to Observed

The 10 km input dataset can be compared to observed in two ways: examining the profile by comparing one model point to the rawinsonde, or examining the surface temperature by comparing all model points to a gridded observed dataset. We will do both.

In Figures 55, 56, 57, and 58, we compare the point in the input dataset closest to SLC to the SLC rawinsonde. The input dataset's temperature is 6 degrees Celsius too warm, and up to 2 g/kg too moist. Winds are within 4 m/sec of observed. Relative humidity is within about 20% of observed, being 10% too low near the surface and 10% too high around 500 hPa. Both temperature and specific humidity are too high, but the increase compensates for each other to give good agreement with relative humidity. The differences in humidity and wind speed are at the observed variability, but temperature is significantly too warm

The RMS difference between the observed monthly surface temperatures over the domain is 1.3 degrees Celsius. In Figure 59 we compare the RMS differences between the input 10 km dataset of surface temperature and the gridded observed surface temperature. The observed temperature dataset is created with observed station temperatures using the LP regression as detailed in chapter II and the error estimates are given in Figure 60. RMS difference errors are roughly 1 degree Celsius higher than the observed LP error estimates. Thus we conclude there is reasonable agreement between the observed and the 50 km input dataset, with the exception that the input dataset is much too moist and warm near SLC.

IV.2.5 Simulation Input Dataset Compared to the ECMWF Dataset

The ECMWF dataset, just validated in section IV.2.4, is now compared to the 10 km simulation input dataset. As shown in Figures 61, 62, 63, and 64, the 50 km dataset is moister than ECMWF. The input dataset's temperature is within 2 degrees of ECMWF and winds are too fast by 4 m/sec. Specific humidity is much too high by 2 mg/kg, with significant differences throughout the entire profile. Relative humidity is also conspicuously higher than observed. The differences in temperature and wind speed are within observed variability. However, humidity is much moister than ECWMF. Since the ECMWF dataset itself is too moist near SLC, this certainly indicates the input dataset is much moister than reality. The input dataset was also shown to be significantly too warm near SLC, which means considerable variation can exist between individual data-points.

In conclusion, we find that the errors in our input dataset are similar to the errors in both the observed dataset and the ECMWF dataset. Thus we surmise that our input dataset should provide reasonable LBC for the high-resolution simulation. The one concern that remains is that the input dataset is certainly much more moist than the observed.

IV.2.6 Theoretical Mechanism for Artificially Enhanced LBC Storms

With 6-hour updates starting at 12:00 am Greenwich Mean Time (GMT), we miss some of the daily extreme events. Accounting for daylight saving time, updates at 12:00 am, 6:00 am, 12:00 pm, and 6:00 pm GMT are updates at 5:00 pm, 11:00 pm, 5:00 am, and 11:00 am local standard time. Thus the linear time interpolation will smooth over the hottest part of the day. And, the LBCs between 11:00 am and 5:00 pm will be too cool, with a maximum difference at about 2:00 pm. Since the interior of the model properly responds to solar heating, the difference in temperature on the interior to the LBC produces a temperature gradient near the LBC. This temperature gradient, if significant, can cause cool air to be artificially advected in from the lateral boundaries. As cool air is advected into the system, it can cause enough cooling to bring the air to saturation and create an artificially enhanced storm event. The reason this does not happen in regional models with larger domains is that the temperature gradient is applied over a larger area. The gradient is applied over the "relaxation" region on the lateral boundaries where the adjustment scheme blends input and model data. For our 50 km run the relaxation region consists of the outer 11 boundary grid-points. This implies that an error in the outer boundary has to be applied over a distance of 550 km (11 cells of 50 km each) before the error changes the gradients near the boundary. An error of 3 degrees Celsius on the boundary produces an implied gradient of 5.5×10^{-6} C/m. For the 10 km grid the relaxation region is only five cells or 50 km. The magnitude of the error in the 10 km run is an order of magnitude larger than the implied error in the 50 km run.

This same temperature error of 3 degrees Celsius in the LBC now produces an implied gradient of 6×10^{-5} C/m. The model will apply this gradient both in horizontal diffusion and advection, but for comparison purposes we will examine horizontal advection only.

Advection is a driving term in the continuity, energy, and momentum equations. It is the gradient in wind speed and the given quanitity examined (such as temperature). Let us calculate the advection of temperature for a simple case. Let us assume a constant wind of 5 m/sec and an implied temperature gradient of 3 degrees Celsius applied over the boundary region. Consequentially, the implied advective cooling will be -3×10^{-4} C/sec. The above is of the same magnitude as the dominant terms in the energy equation are. For the 50 km simulation this term will be less than a tenth that size and therefore will not be a dominant term. Also, since a significant part of the LBC for the 50 km domain is over ocean, an error of 3 degrees Celsius is less likely than in the 10 km domain where an error of that magnitude is very common. This is because oceans experience an insignificant diurnal variation compared to landmasses. Thus errors in time interpolations are much smaller over the ocean.

In equations 1 and 2 we show the formula for water vapor saturation. When we use the values for 700 hPa and 280 K we get a saturation-mixing ratio of 0.90 g/kg. With a 3-degree cooling to 277 K, we get saturation at 0.73 g/kg. Thus the 3-degree error artificially changes 80% relative humidity to 100% relative humidity.

The formula for water vapor saturation is:

$$q_{SAT} = 0.622 p_{SAT} / (p - 0.378 p_{SAT}) \quad (1)$$

$$p_{SAT} = 6.11 \exp(19.84659 - 5418.12/T) \text{ (hPa)} \quad (2)$$

where psat in equation 2 is the saturation pressure (hPa); p is the pressure (hPa); qsat is the water vapor mixing ratio at saturation (kg/kg); and T is the temperature (K).

It is quite plausible that as this sort of error gets advected through the system, it can artificially enhance a storm moving through the domain. As we see in the results section, spurious storms do not seem to be created through this mechanism, but it can alter a storm system as it comes into the model domain.

IV.3 RESULTS

In the discussion we establish that the input dataset to our 10 km simulation compared well to observed. Although there are some concerns with the dataset (see section IV.2.2), the input fields are near to the observed values and should provide good input LBCs.

In section 1 we introduced two problems with our 10 km simulation. The first was the general lack of precipitation over the entire grid-cell, and the second was the overabundance of precipitation associated with mountain up-slopes.

We believe the first problem (lack of precipitation) is due to two problems. The first and primary problem has to do the cloud microphysics. The explicit cloud parameterization creates thick clouds that do not produce significant precipitation. The cloud parameterization (explicitly the auto-conversion formulation) needs to be changed so that more of the cloud water is converted into precipitation. The second problem has to do with our lack of providing cloud and rain water as input at the LBC. This causes two changes. One, precipitation will fall near the lateral boundaries, a region almost completely dry currently. Secondly, clouds will be "spun-up" to precipitating levels faster, since they enter the model domain as cloud rather than having to develop within the small model domain.

The second problem (abnormally high amounts of precipitation on up-slopes) was due to the "artificially enhanced LBC storms" that were discussed in section IV.2.6. A closer examination of these two problems is warranted.

Let us first turn our attention to the general lack of precipitation in the model domain. In Figure 65 we compare the total precipitable water of the input dataset to ECMWF over the 10 km simulation domain. Precipitable water is the total water mass in the atmosphere. Total mass over the domain of the observed precipitation is 3.08×10^{12} kg while the same value for the 50 km and 10 km simulations are 4.24×10^{12} kg and 2.09×10^{11} kg, respectively. Thus total precipitable water is an order of magnitude higher than the precipitation. Likewise, cloud water is an order of magnitude higher than precipitation as shown in Figure 66. The input 50 km simulation is a little high, as we can see from these values. However, the resulting 10 km simulation is an order of magnitude lower than the observed and 50 km simulation. See Figure 67 for a plot of the average precipitation over the domain area for each day. The input precipitable water as shown in Figure 65 is slightly high, which confirms the excess precipitation in the 50 km simulation, but disagrees with the low level of precipitation in the 10 km simulation. We must consider the fact that the

same conditions appear for the 50 km simulation as well. Values that are interpolated from the 50 km simulation drive the 10 km simulation. So why does the 10 km simulation receive so much less precipitation than the 50 km simulation if the two have similar conditions?

The bias in precipitable water leads us to consider possible problems with the interpolation of the 50 km simulation for input into the 10 km simulation. A concern with the interpolation of the 50 km dataset is that the interpolation was done on sigma levels rather than pressure surfaces. Interpolation on sigma surfaces mixes the horizontal and vertical derivatives and impacts the end result. Extrapolating temperatures for 10 km terrain beneath the 50 km surface is also problematic. *Trenberth et al.* [1993] dealt with these issues in detail. We implement their procedures for vertical interpolation. This amounts to interpolating the 50 km surface, first to pressure levels, then horizontally to the 10 km grid, and, at last, vertically interpolating the pressure levels to the final 10 km sigma coordinates. We also use the *Trenberth et al.* [1993] procedures to extrapolate temperatures below ground. Our treatment of humidity also differs. Because water vapor mixing ratio is conserved in the boundary layer, we interpolate on the mixing ratio near the surface rather than relative humidity. At higher levels we interpolate on the relative humidity as in *Trenberth et al.* [1993]. We also use a simple log profile to compensate for the winds that became closer to the ground after the interpolation. Figures 68, 69, 70, and 71 show the difference between our new LBC dataset and the original control simulation. The differences are negligible. The total water mass is unchanged between the two datasets. And the differences in other variables are very small and much less than the variability in the datasets.

There is another source in the water budget that we have not considered. That source is cloud and rain water. The 50 km simulation handles precipitation in a different manner than the 10 km simulation. The 50 km simulation uses implicit cloud physics, meaning that cloud water is diagnosed rather than prognostically solved for. The 10 km simulation uses explicitly prognosed cloud and rain water. Since the 50 km simulation did not have cloud and rain water, we did not include it as part of the LBCs. Physically this implies that no cloud or rain water gets advected into the system from outside the system. Anyone familiar with a desert area realizes that the majority of the water comes from storms that have developed outside the area. The implicit form of precipitation in the 50 km simulation means that precipitation immediately appears on the surface as rain. Thus, there is no time for this rain water to be advected into the 10 km domain. Were the interface between the 50 km and 10 km simulations realistic, the cloud and rain water from the 50 km domain would be advected into the 10 km domain. Since the 50 km simulation was done without cloud and rain water, we can only examine the total amount of water that would be advected into the system from the 50 km domain. To estimate this we total the amount of rain over the 50 km cells on the 10 km LBC. To make sure that we only get the rain that would be advected into the system, we only sum those grid-points that have a wind moving into the 10 km domain. We estimate that up to 1×10^{12} kg of rain water would be available for advection into the 10 km simulation in this manner. This raises the precipitation bias to about half of the observed. Most of this water will be deposited near the lateral boundaries and not impact the total simulation. However, some portion will enter the system as cloud water and affect the interior solution as the source is driven into the simulation domain. However, since this source of water only accounts for a fraction of the precipitation deficit, there must be other sources of water to consider.

Two other sources of additional water are the cloud water in the system, and water that can be released from the cumulus parameterization. We will consider cloud water first. The total column

cloud water totaled over the domain is 74 mm/m^2. As such, there is sufficient water that could be converted into precipitation while still allowing clouds in the system. See Figure 66 for the distribution of total column cloud water for the entire period through the model domain. Notice that cloud water is fairly evenly distributed, so converting more cloud water into rain water will create a water source evenly distributed over the domain. Since cloud water forms as water vapor is saturated, it is unlikely that the precipitation deficit is because of supersaturation of water vapor. However, the cloud water physics could be tuned to allow for more precipitation. The most important term is the auto-conversion term that initiates the creation of rain droplets. RegCM2 uses a *Kessler* [1969] expression for auto-conversion as follows:

$$\mathbf{P_{RC} = k_1(q_c - q_{cth})} \qquad \mathbf{(3)}$$

where q_{cth} = 0.41 g/kg, $k_1 = 1x10^{-3}$ 1/m, qc is the cloud water mixing ratio, and P_{RC} acts as a source term for rain water and a loss term for cloud water.

Anthes et al. [1987] mention that this parameterization is sensitive to resolution, and *Hsie* [1983] did some limited sensitivity studies. Thus, it is not surprising that some tuning may be in order for a different resolution. There is a substantial amount of cloud water above the threshold of 0.3 g/kg (from Figure 66). Cloud water is also spread out over the entire domain area, which should improve the grid-averaged statistics. Lowering the auto-conversion threshold will increase precipitation. In experiments at 20 km resolution over a similar domain for July 1981, we found lowering the threshold from 0.5 g/kg to zero doubled the maximum precipitation amounts. Assuming total precipitation doubles, this still brings us quite short of the precipitation deficit for our situation. So although sufficient water is available in the clouds to account for the precipitation deficit, realistic changes to the auto-conversion term do not allow large enough increases in precipitation. Lowering the auto-conversion threshold is akin to allowing precipitation to form in clouds less "thick" with moisture. As we discussed in section 2.3, the model is at an awkward resolution. The resolution is high enough that most of the convective activity is resolved. However, the grid size is large enough that the vertical velocities are likely to be much lower than might be observed in an actual cumulonimbus cloud system. Therefore, the model does not produce as much condensation as might have been produced in the cumulus parameterization. Hence, the condensation that it does produce must be appropriately allocated. Allowing clouds to form at saturation levels less than 100% may be required to account for the precipitation deficit.

The next water source to consider is the possibility of getting more precipitation out of the cumulus parameterization. However, first we consider this likelihood. The cumulus parameterization accounted for approximately 35% of the precipitation for the 50 km simulation over this region for September 1981. At a higher resolution we expect that much of the activity dealt with in the larger-scale simulation will be resolved at the 10 km level, and we expect a reduction in the percentage of precipitation coming from a subscale source. Even the same percentage as the 50 km simulation (35%) would not explain the precipitation deficit. What was observed was approximately 0.5% of the amount in the 50 km simulation. Increasing this percentage to even 5% may take substantial changes to the parameterization. And in the end, these changes seem unlikely to affect the overall quality of the model. Thus, it seems the best candidate for tuning changes is the auto-conversion parameterization. One possibility is to replace the auto-conversion parameterization with a more modern expression that is less sensitive to changes in resolution. We may also need to allow cloud water to form at subsaturated levels.

The second problem we wish to discuss is artificially enhanced LBC storms. As a first check for LBC storms we run the model at a higher LBC update and observe if the simulation changes significantly. If the update frequency is sufficiently high, we expect almost no difference between the two solutions. In Figure 72 we show a simulation with hourly LBC updates. The precipitation distribution has changed, but the amounts are much less than the control simulation with 6 hour LBC updates. We need to examine the simulation in more detail to understand how and why the simulation changed. Figure 73 shows the difference between the grid-scale-averaged precipitation for the control to the 1-hour update. There is a certain amount of low-level noise that is expected but there are seven spikes of significant magnitude in the series. Of the seven, six show increased precipitation in the control run and only one shows decreased precipitation. Most of these spikes also initiate at the 6 hour period between 11:00 am and 5:00 pm, and peak this same time as well. So although this evidence is circumstantial, it seems to indicate artificial enhancements of several of these storms. However, at the same time any statistical measure of performance that we apply to either simulation, specifically the RMS differences of precipitation with observed and zero level threshold threat scores, shows that neither simulation is preferred. But, the reason for this is the model's unreasonable performance at this resolution and entirely too low precipitation amounts. Once the changes to the explicit cloud water formulation are made, this experiment needs to be performed again.

IV.4 DISCUSSION AND CONCLUSIONS

We have discussed some of the difficulties associated with modeling precipitation in LAM at the 10 km scale. Our simulations exhibited the same behavior shown by other simulations – not enough precipitation over the grid domain coupled with too much precipitation over the mountain up-slopes. We have done extensive analysis on these simulations and concluded the following. Cloud and rain water inputs are important to include on the LBC. And, LBC update frequencies and total domain area impact the final solution.

When we examined the total water budget, we found the precipitable water to be too low in the 10 km system. However, as the same bias was displayed in the 50 km simulation, we felt the answer lay elsewhere. One bias is the lack of input cloud and rain water into the system. This bias may account for 50% of the total deficit in precipitation. We believe the rest of the bias has to do with the error in precipitable water already discussed, as well as the lower than observed variability in relative humidity. At this point some tuning to the explicit cloud and rain water physics can be done. Although initially we did preliminary experimentation with the cloud water physics, our results were unsuccessful. The total bias was too large for a single change to one parameter to affect the results. By doing extensive analysis and attacking the problem on several fronts, we think we have arrived at a successful solution to the problem.

Now we consider concentration of rainfall on windward up-slopes. Much of this problem has to do with the grid-scale average problem. Since precipitation does not occur en masse over the grid domain, the only precipitation that does form is via orographic uplift. By allowing cloud and rain water to be advected in from the LBC, more precipitation will form over the entire grid domain.

As to the point brought up by *Kato* [1997] in regard to "determining the optimal dynamical framework," our contribution seems to be increased understanding of one method – one-way nesting. There are some problems unique to this setup that are exasperated at high resolutions. Sufficiency of domain size and LBC update frequency is an important issue for one-way nesting.

One-way nesting is a popular method of incorporating large-scale information into a LAM. It is likely that it will remain so. Even models with two-way nesting require one-way nesting with large-scale input fields. As GCM resolutions increase, the technology to one-way nest LAMs to very high resolution GCM simulations will become important, hence the utility of our study.

Kato [1997] continues with the question of "the optimal precipitating scheme." We found that explicit cloud and rain water were important components of these 10 km simulations. Our contribution is to point out the need for explicit precipitation physics in high-resolution LAM simulations. We also point out that for our domain, convective paramaterization did not play a role.

As to the question of the utility of high resolution climate simulations – we can only say this issue warrants further study. Observation networks do not exist at resolutions required to establish the fine details of these simulations. And, there are significant tuning problems to be worked out with such simulations.

Finally, can LAMs be used to provide accurate answers to regional concerns of climate change and run impact models? Once the model is sufficiently tuned we believe so. However, the one-way nesting to 50 km and 10 km simulations is limited. This is due to the fact that the biases of the GCM are input into the 50 km and 10 km models. This is compounded with the biases of the 50 km and 10 km models and affects the final 10 km simulation. This is another reason the accurate analysis of the model simulations must be done at each step. It is also important to understand the biases inherent in going between each model simulation, gaining an understanding of the total bias of the system. This is why it is vital to compare results for observed sequences, such as we have done here, in order to gain accurate understanding of the total error in the system.

References

Anthes, R. A., Hsie E.-Y., Kuo Y.-H., Description of the Penn State/NCAR mesoscale model version 4 (MM4), *NCAR Tech. Note, TN-282+STR*, 66 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1987.

Briegleb, B. P., Delta-Eddington approximation for solar radiation in the NCAR Community Climate Model, *J. Geophys. Res.*, *97*, 7603-7612, 1992.

Cotton W. R., and R. A. Anthes, *Storm and Cloud Dynamics*, 883 pp., Academic Press, San Diego, Calif., 1989.

Davies, H. C. and R. E. Turner, Updating prediction models by dynamical relaxation: An examination of the technique, *Q. J. R. Meteorol. Soc.*, *103*, 225-245, 1977.

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, Biosphere-atmosphere transfer scheme (BATS) version 1E as coupled to the NCAR community climate model, *NCAR/TN-387+STR*, 69 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1993.

Fritsch, J. M., and C. F. Chappell, Numerical prediction of convectinvely driven mesoscale pressure systems. Part I: Convective paramertization., *J. Atmos. Sci., 37*, 1722-1733, 1980.

Giorgi, F., Simulation of regional climate using a limited area model nested in a general circulation model, *J. Clim.*, *3*, 941-963, 1990.

Giorgi, F., Sensitivity of simulated summertime precipitation over the western United States to different physics parameterizations, *Bull. Am. Meteorol. Soc., 119*, 2870-2888, 1991.

Giorgi, F., and G. T. Bates, The climatological skill of a regional model over complex terrain, *Mon. Weather Rev., 117*, 2325-2347, 1989.

Giorgi, F., G. T. Bates, and S. J. Nieman, Simulation of the arid climate of the Southern Great Basin using a regional climate model. *Bull. Amer. Meteorol. Soc., 73*, 1807-1822, 1992.

Giorgi, F., G. T. Bates, and S. J. Nieman, The multiyear surface climatology of a regional atmospheric model over the western United States, *J. Clim., 6*, 76-95, 1993a.

Giorgi, F., M. R. Marinucci, and G. T. Bates, Development of a second-generation regional climate model (RegCM2). Part I: Boundary-layer and radiative transfer processes, *Mon. Weather Rev., 121*, 2794-2813, 1993b.

Grell, G. A., J. Dudhia, and D. R. Stauffer, A description of the fifth generation Penn State/NCAR mesoscale model (MM5), NCAR Tech. Note, NCAR/TN-398+STR, 121pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1994.

Holtslag, A. A. M., E. I. F. de Bruijin, and H. L. Pan, A high-resolution airmass transformation model for short-range weather forecasting, *Mon. Weather Rev., 118*, 1561-1575, 1990.

Holtslag, A. A. M., and B. A. Boville, Local versus non local boundary-layer diffusion in a global climate model, *J. Clim., 6*, 1825-1842, 1992.

Hong, S.-Y., and H.-M. H. Juang, Orography blending in the lateral boundary of a regional model, *Mon. Weather Rev., 126*, 1714-1718, 1998.

Hostetler, S. W., and F. Giorgi, Use of a regional atmospheric model to simulate lake-atmosphere feedbacks associated with Pleistocene lakes Lahontan and Bonneville, *Clim. Dyn., 7*, 39-44, 1992.

Hsie, E.-Y., Frontogenesis in a moist atmosphere, Ph.D. Thesis, Department of Meteorology, Pennsylvania State University, Penn, 251 pp., 1983.

Hsie, E.-Y., and R. A. Anthes, Simulations of frontogenesis in a moist atmosphere using alternative parameterizations of condensation and precipitation. *J. Atmos. Sci., 41*, 2701-2716, 1984.

Kato, T., Hydrostatic and non-hydrostatic simulations of moist convection: review and further study, *Meteorol. Atmos. Phys., 63*, 39-51, 1997.

Kessler, E., III, On the distribution and continuity of water substance in atmospheric circulation's, *Meteorol. Monogr., AMS, No. 27*, 84 pp., 1969.

Lall, U., and K. Bosworth, Nonparametric statistical inference and function estimation for hydrologic space time data, *Trends Hydrol.*, CSRI, 1-19, 1994.

Madala, R. V., Efficient time integration schemes for atmosphere and ocean models, in *Infinite-Difference Techniques for Vectorized Fluid Dynamics Calculations*, edited by D. L. Book, pp. 56-74, Springer-Verlag Publ., New York, 1981.

Matthews, D. M., F. Giorgi, and G. T. Bates, Regional model simulation of an extreme precipitation event in the Colorado river basin, in *Proceedings of the Third Symposium on Global Change*, AMS, Boston, Mass., 20-23, 1992.

Medina, J. G., Application of a simple local-scale numerical model in the study of altered climate impacts on watershed precipitation, in *Proceedings of the Second Symposium on Global Change Studies*, AMS, New Orleans, La, 68-73, 1991.

Molinari, J., and M. Dudek, Parameterization of convective precipitation in mesoscale numerical models: A critical review, *Mon. Weather Rev.*, *120*, 326-344, 1992.

Nickerson, E. C., E. Richard, R. Rosset, and D. R. Smith, The numerical simulation of clouds, rain, and airflow over the Vosges and Black Forest mountains: A meso-beta model with parameterized microphysics, *Mon. Weather Rev.*, *114*, 398-414, 1986.

Paoli, G. M., *Climate Change, Uncertainty and Decision-Making*, 164 pp., Institute for Risk Research, University of Waterloo, Waterloo Ontario, Canada, 1994.

Sass, B. H., and J. H. Christensen, A simple framework for testing the quality of atmospheric limited-area models, *Mon. Weather Rev.*, *123*, 444-459, 1995.

Trenberth, K. E., J. C. Berry, and L. E. Buja, Vertical interpolation and truncation of model-coordinate data, *NCAR Tech. Note.*, *TN-396+STR*, 54 pp., National Center for Atmospheric Research, Boulder, Colo., USA, 1993.

Waldron K. M., J. Paegle, and J. D. Horel, Sensitivity of a spectrally filtered and nudged limited-area model to outer model options, *Mon. Weather Rev.*, *124*, 529-547, 1996.

Zhang, D., E. Hsie, and M. W. Moncrieff, A comparison of explicit and implicit predictions of convective and stratiform precipitating weather systems with a meso-Beta-scale numerical model, *Q. J. R. Meteorol. Soc.*, 114, 479, 31-60, 1988.

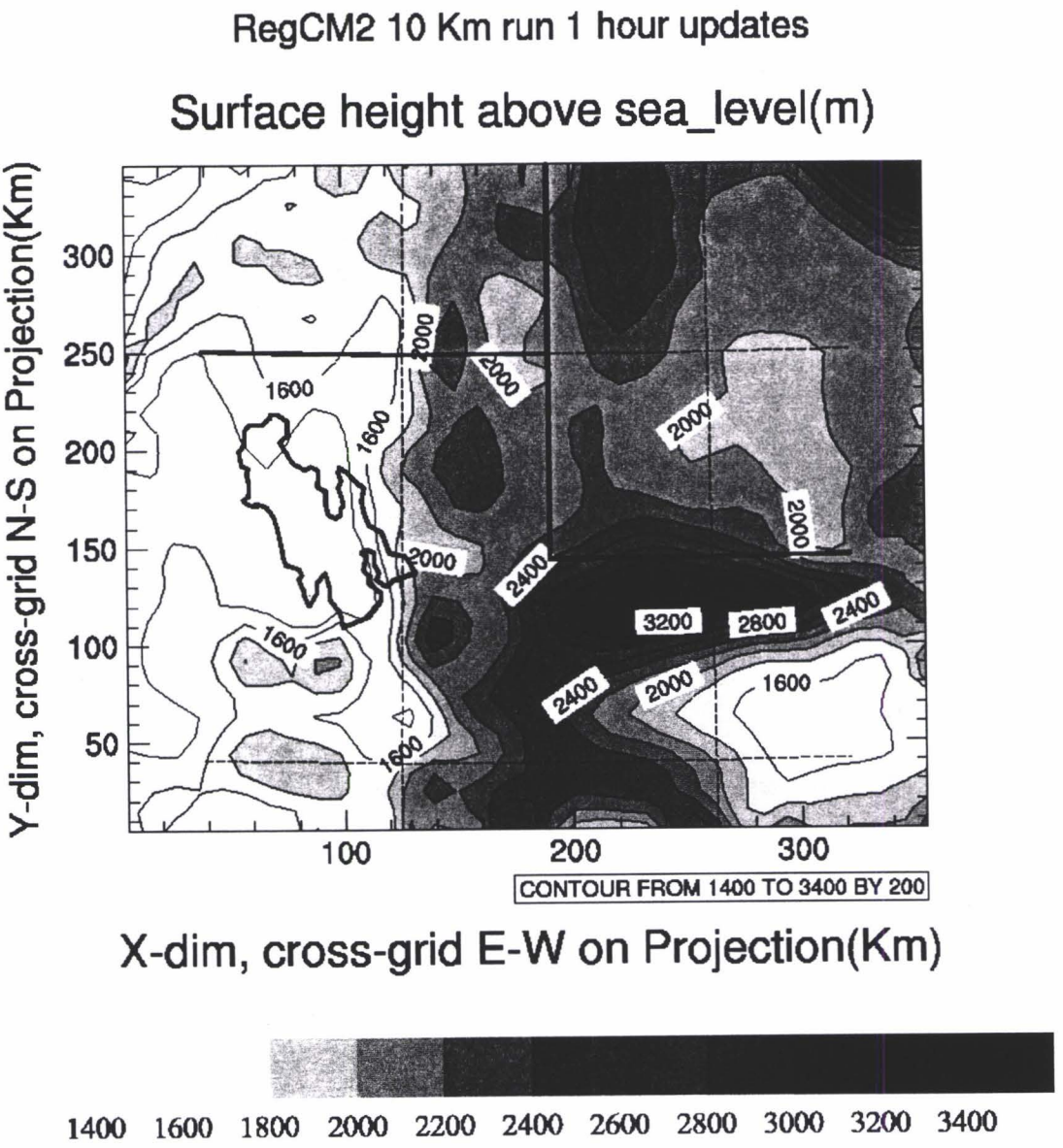

Figure 46. Surface heights for the 10-km model domain.

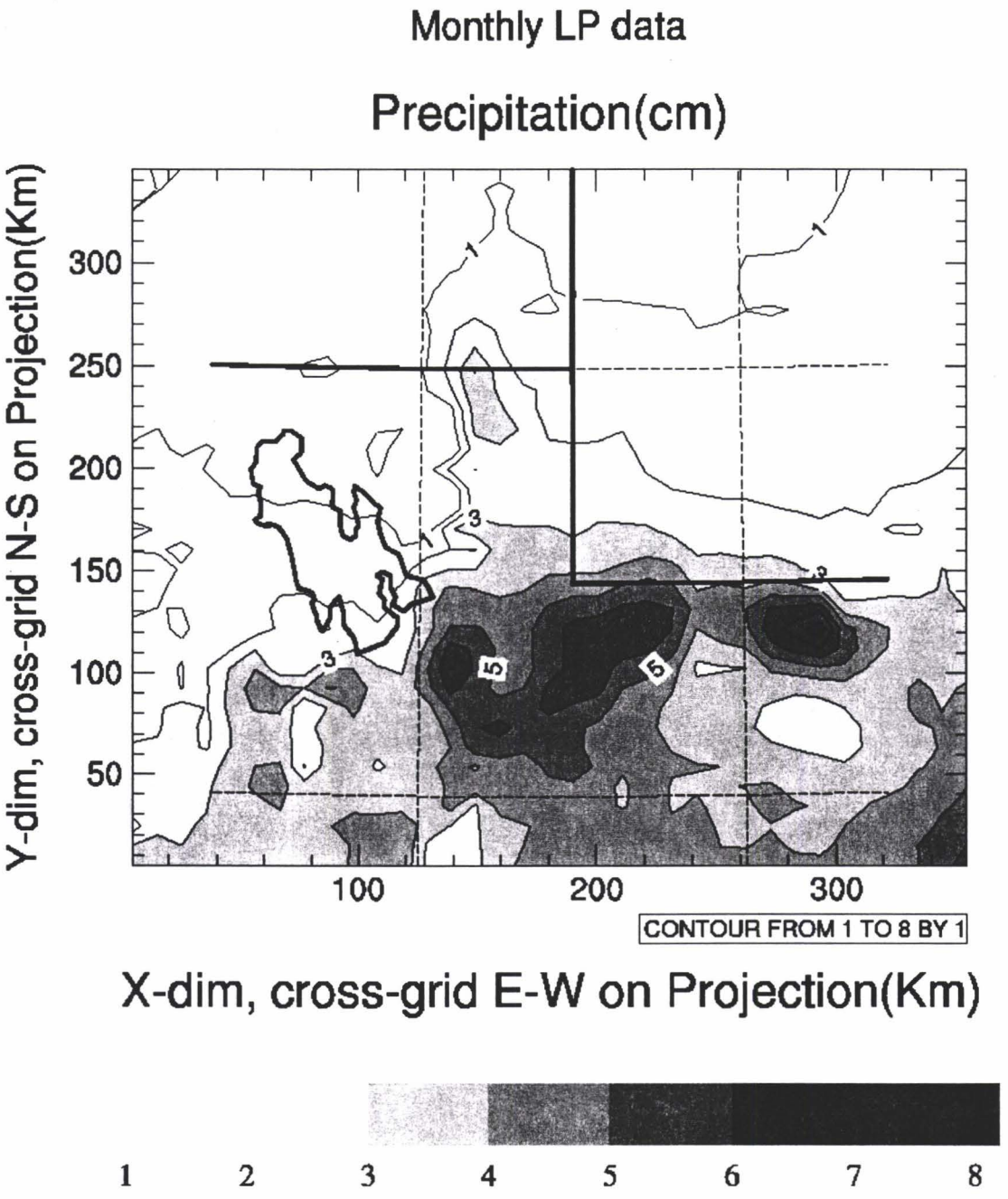

Figure 47. Observed precipitation over model domain for September 1981. Monthly station data was gridded to the model grid using the local polynomial regression technique.

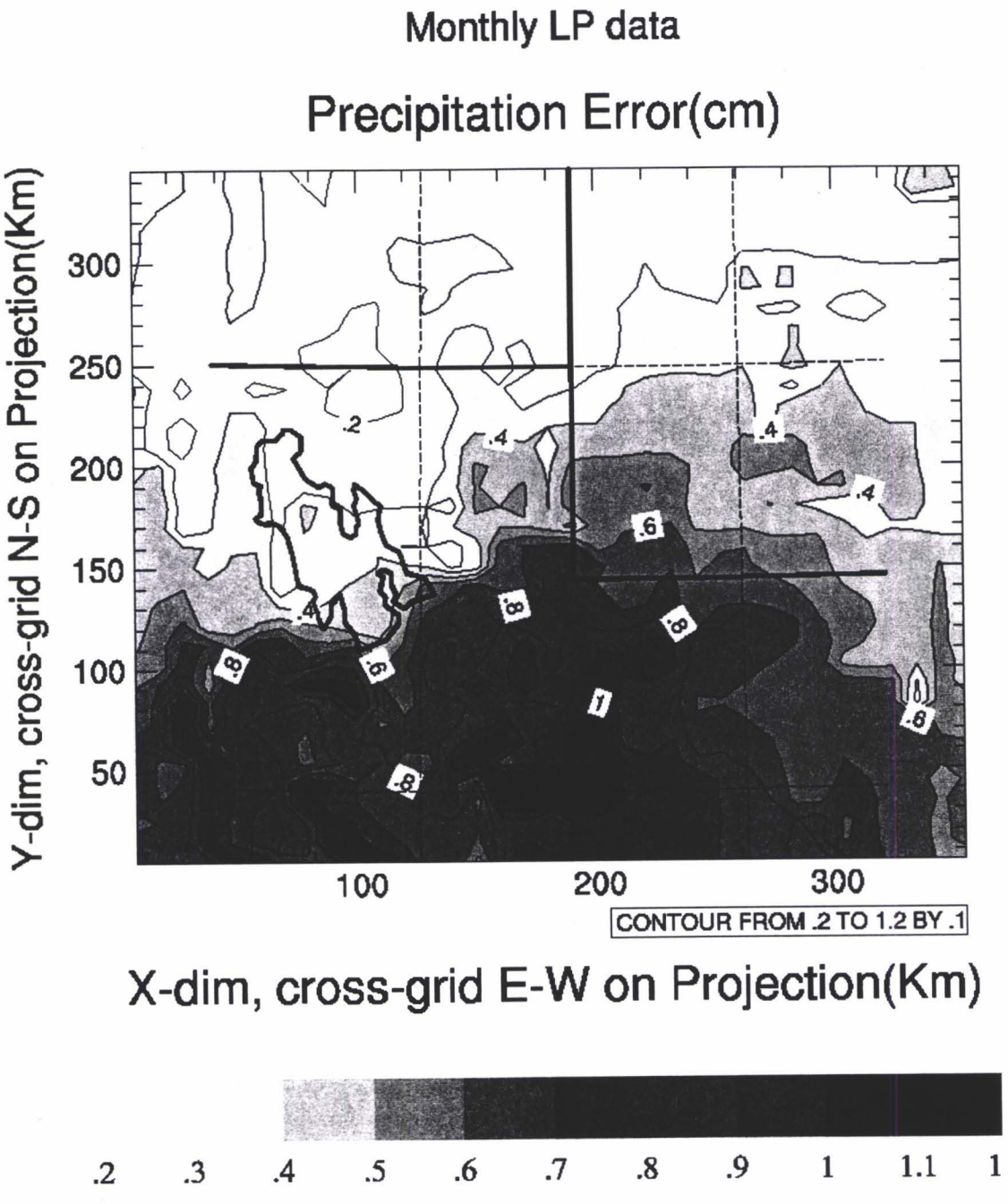

Figure 48. Observed precipitation error over model domain for September 1981. Observed errors are calculated using the local polynomial regression technique.

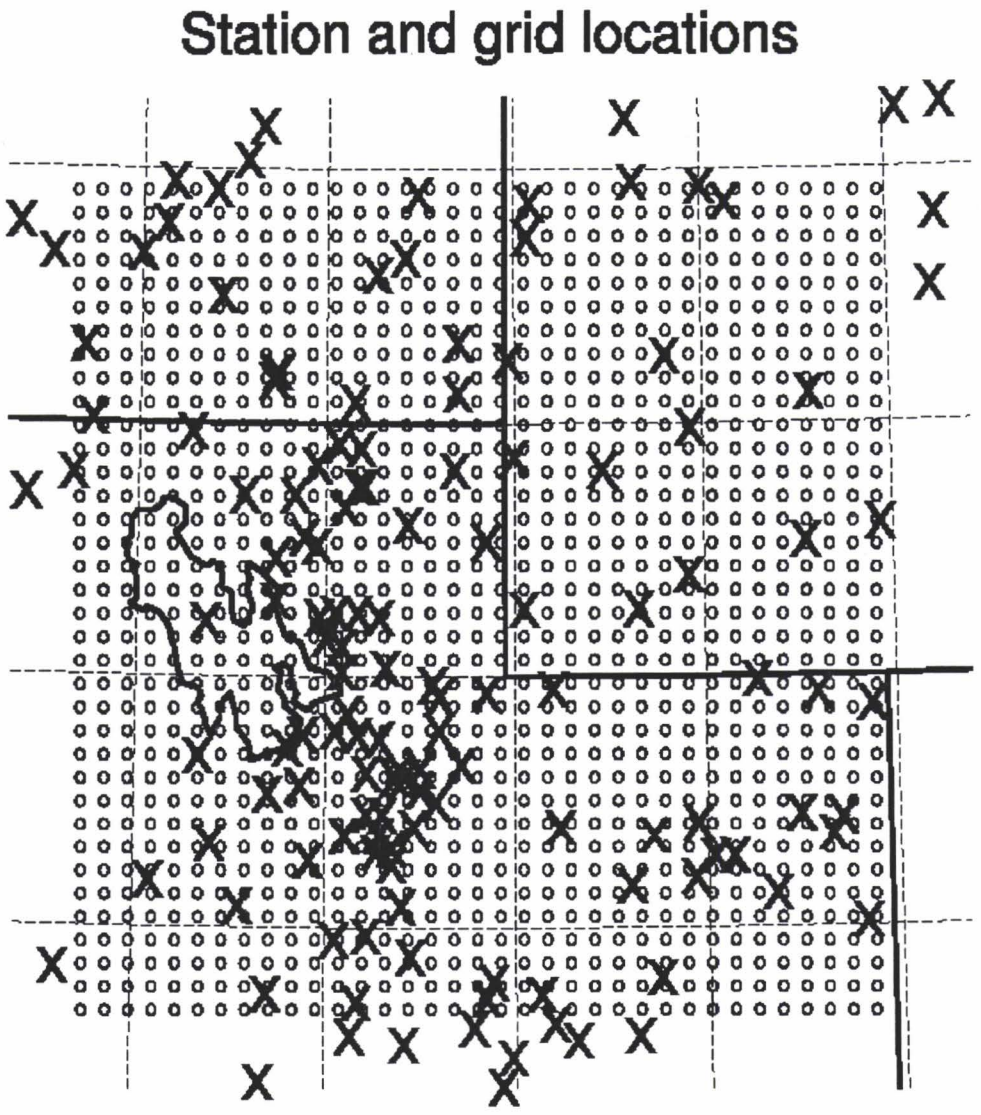

Figure 49. Locations of the observed stations and model grid points for the 10 km domain. Grid points are the green circles and the stations are the red x's.

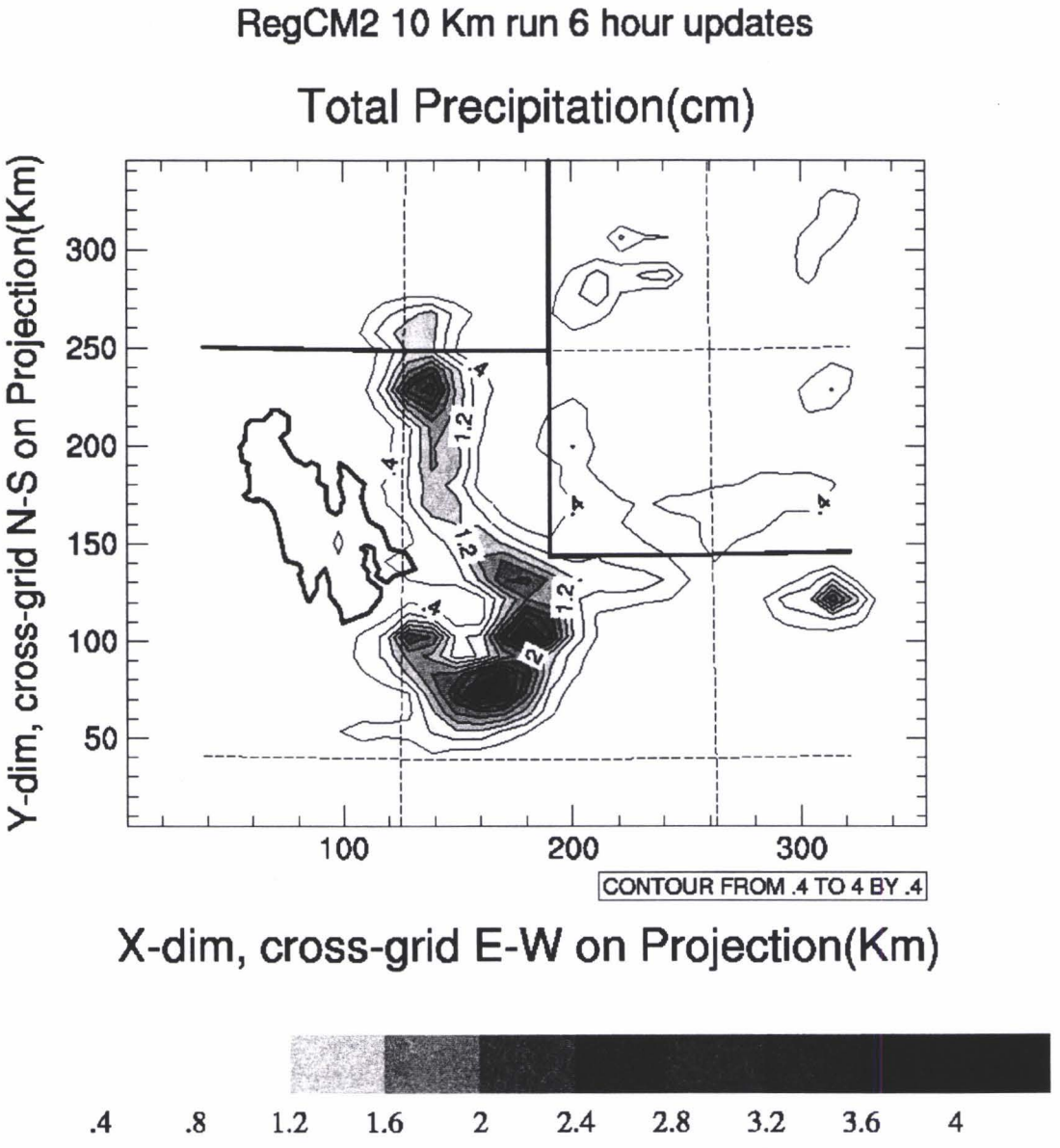

Figure 50. Accumulated precipitation from the RegCM2 model control simulation for September 1981.

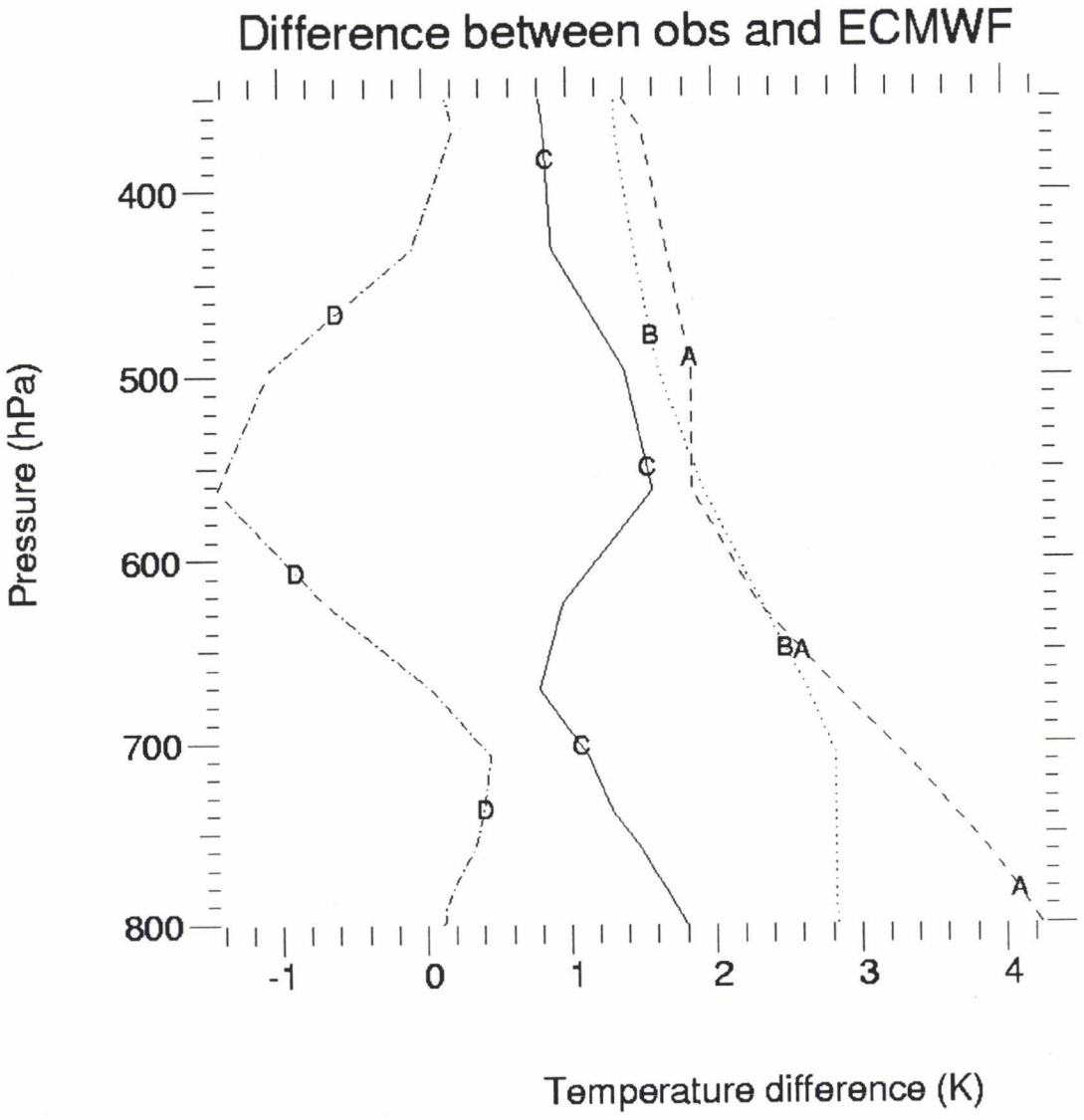

Figure 51. Average and RMS differences (SLC - ECMWF) from ECMWF dataset to SLC rawinsonde, as well as standard deviations for temperature. The ECMWF data was first interpolated to the 10km model domain and then the grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" ECMWF dataset standard deviation. "C", RMS difference between ECMWF and SLC. "D", average difference between ECMWF and SLC.

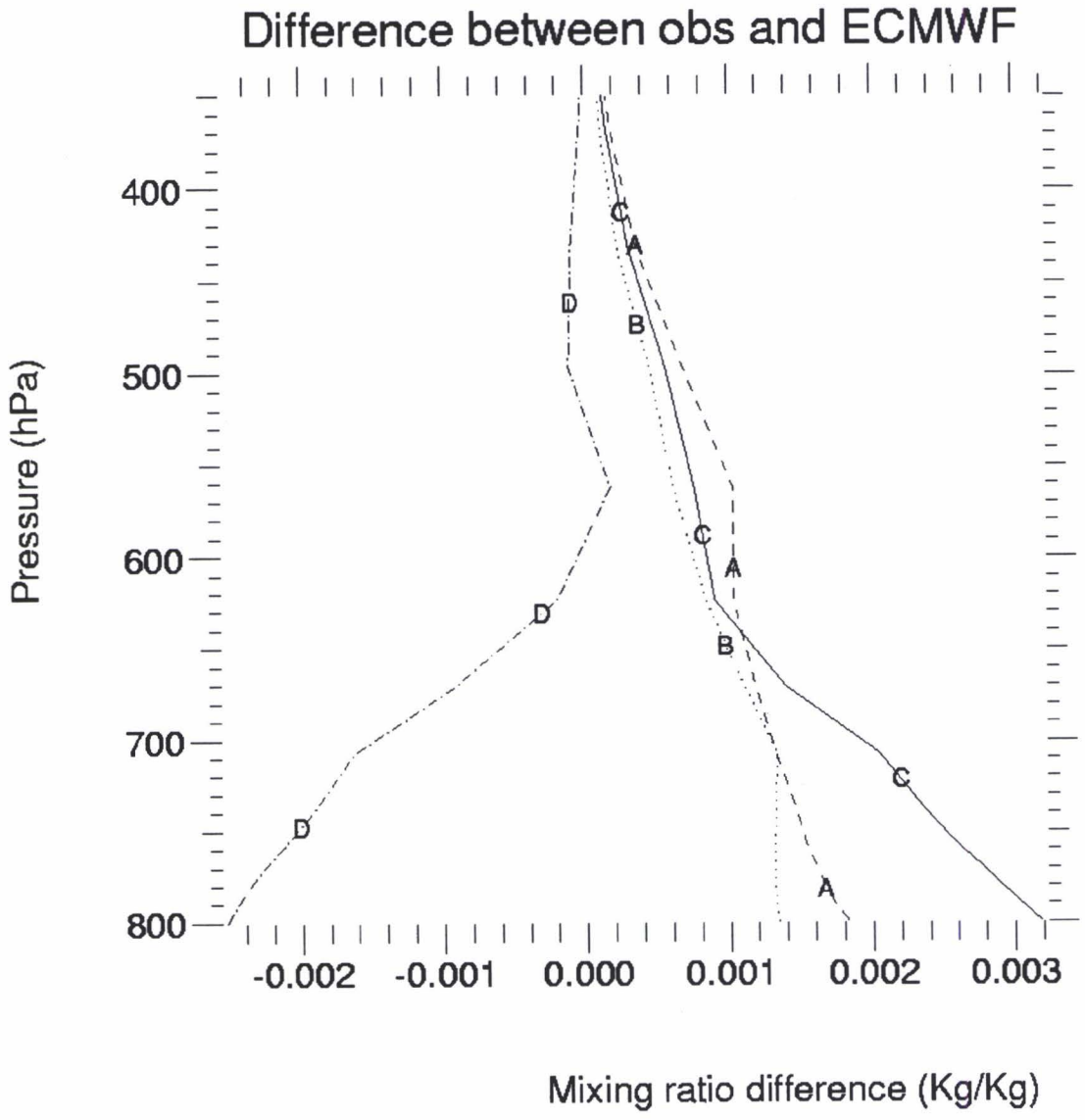

Figure 52. Average and RMS differences (SLC - ECMWF) from ECMWF dataset to SLC rawinsonde, as well as standard deviations for water vapor mixing ratio. The ECMWF data was first interpolated to the 10km model domain and then the grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" ECMWF dataset standard deviation. "C", RMS difference between ECMWF and SLC. "D", average difference between ECMWF and SLC.

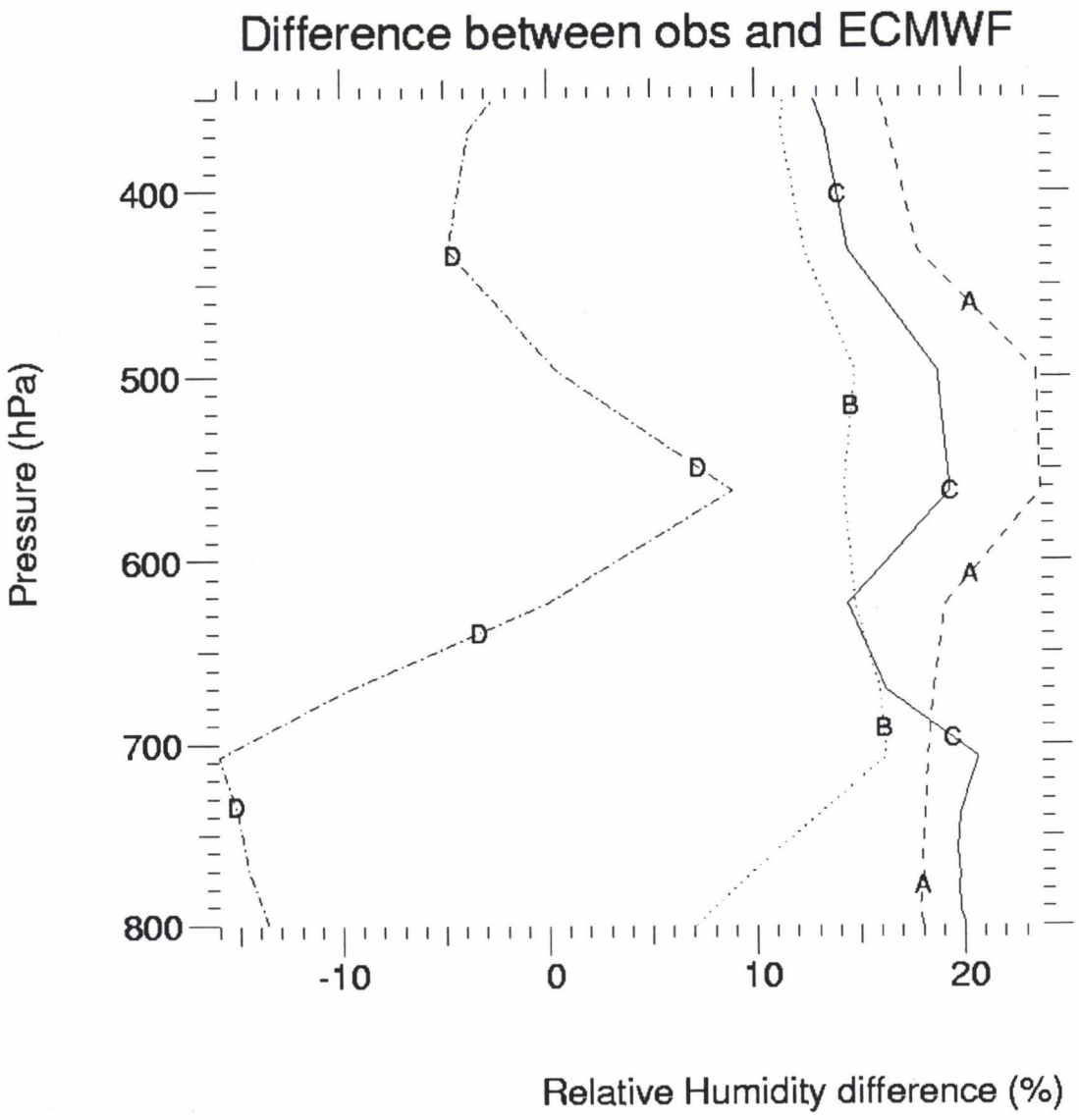

Figure 53. Average and RMS differences (SLC - ECMWF) from ECMWF dataset to SLC rawinsonde, as well as standard deviations for relative humidity. The ECMWF data was first interpolated to the 10km model domain and then the grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" ECMWF dataset standard deviation. "C", RMS difference between ECMWF and SLC. "D", average difference between ECMWF and SLC.

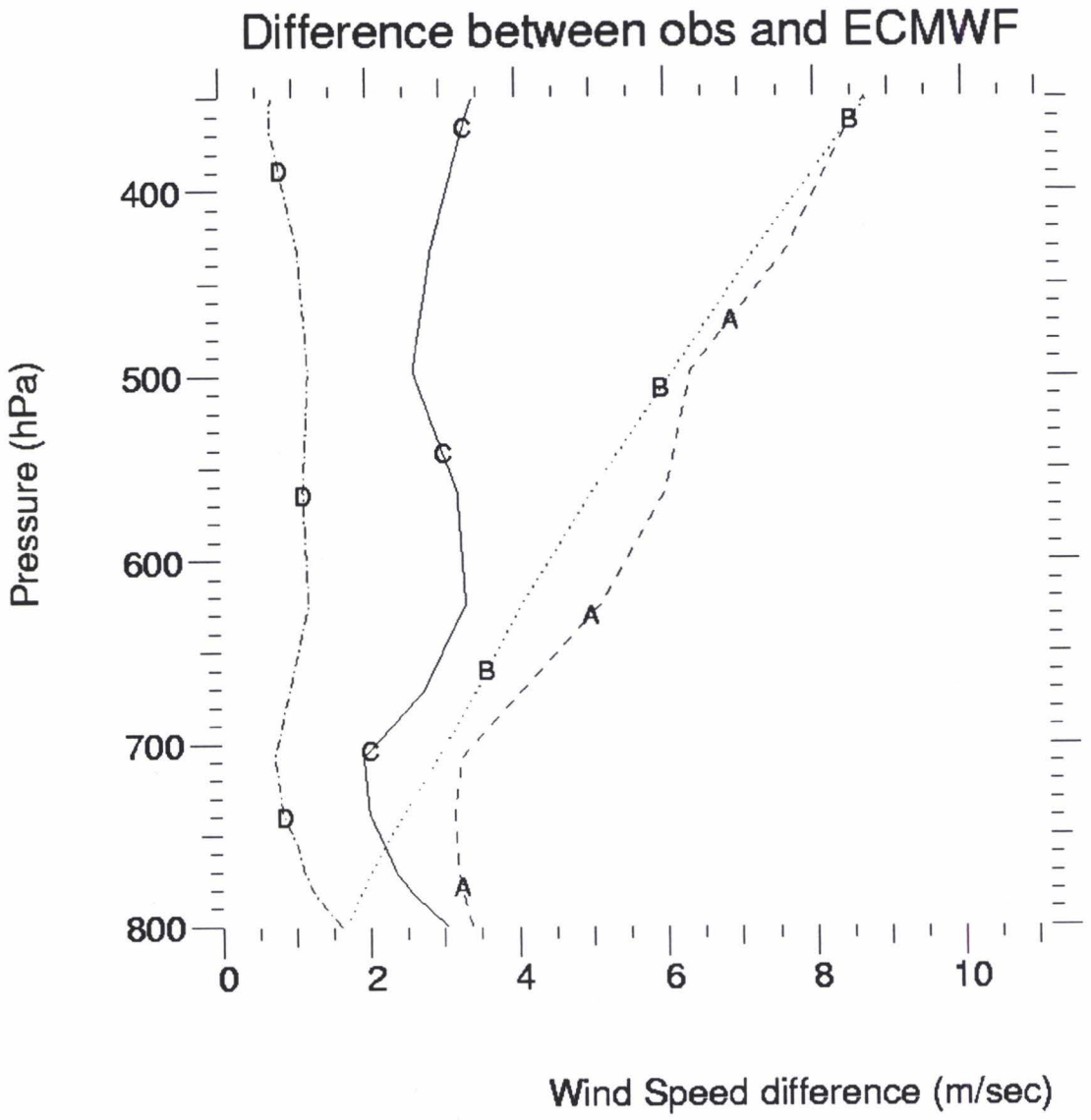

Figure 54. Average and RMS differences (SLC - ECMWF) from ECMWF dataset to SLC rawinsonde, as well as standard deviations for wind speed. The ECMWF data was first interpolated to the 10km model domain and then the grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" ECMWF dataset standard deviation. "C", RMS difference between ECMWF and SLC. "D", average difference between ECMWF and SLC.

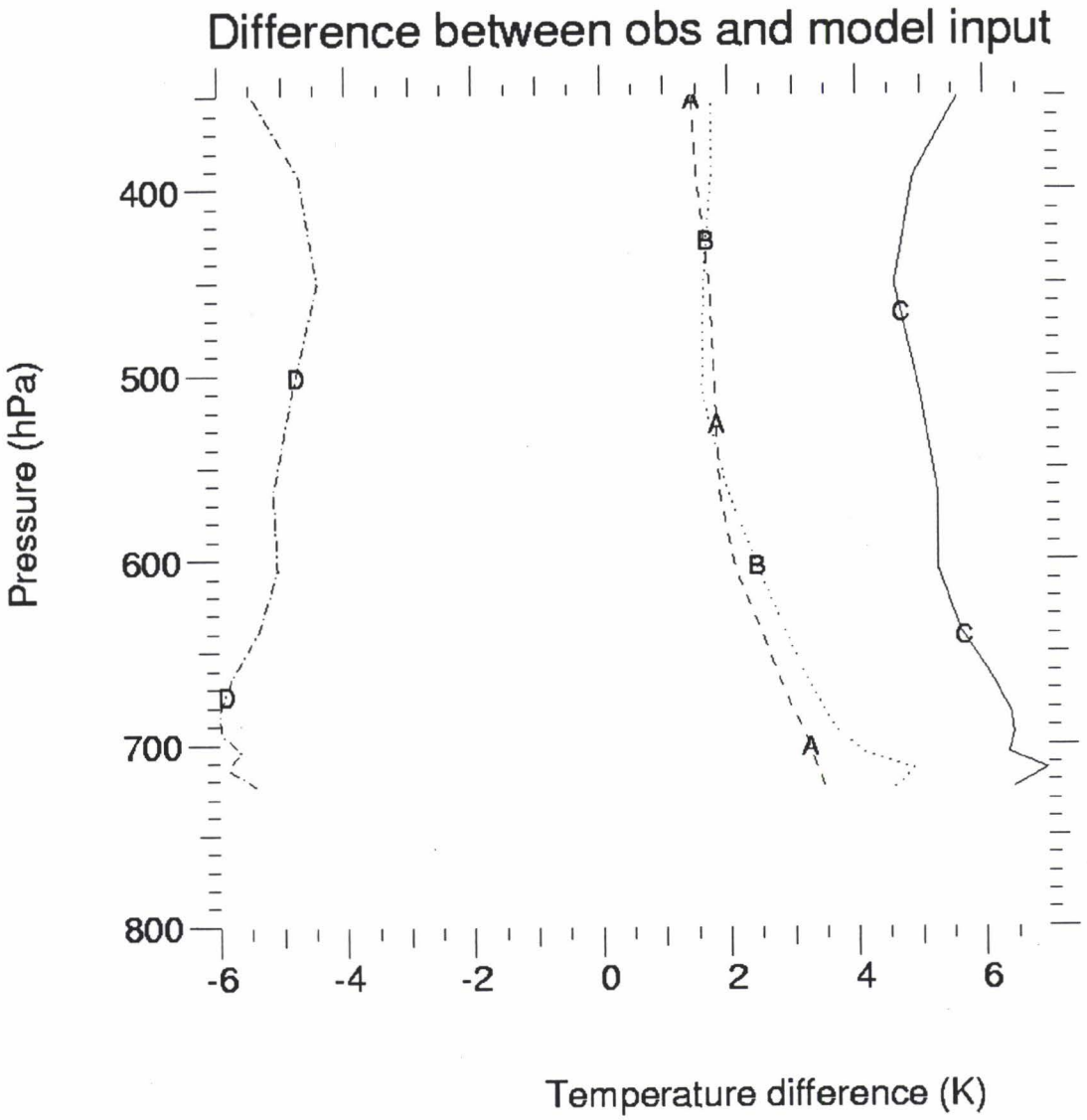

Figure 55. Average and RMS differences (SLC - input) from input dataset to SLC rawinsonde, as well as standard deviations for temperature. The grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and SLC. "D", average difference between input and SLC.

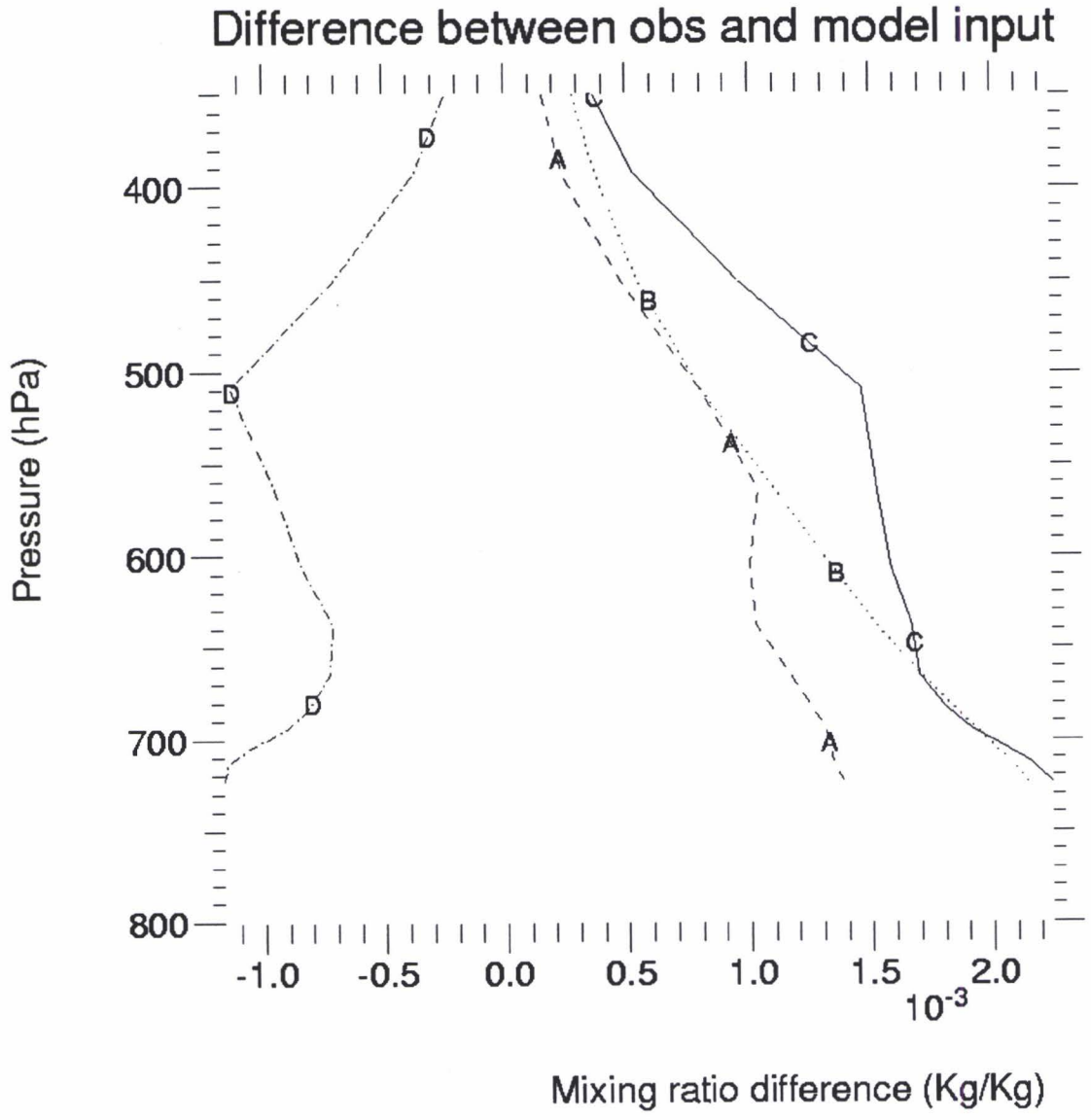

Figure 56. Average and RMS differences (SLC - input) from input dataset to SLC rawinsonde, as well as standard deviations for water vapor mixing ratio. The grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and SLC. "D", average difference between input and SLC.

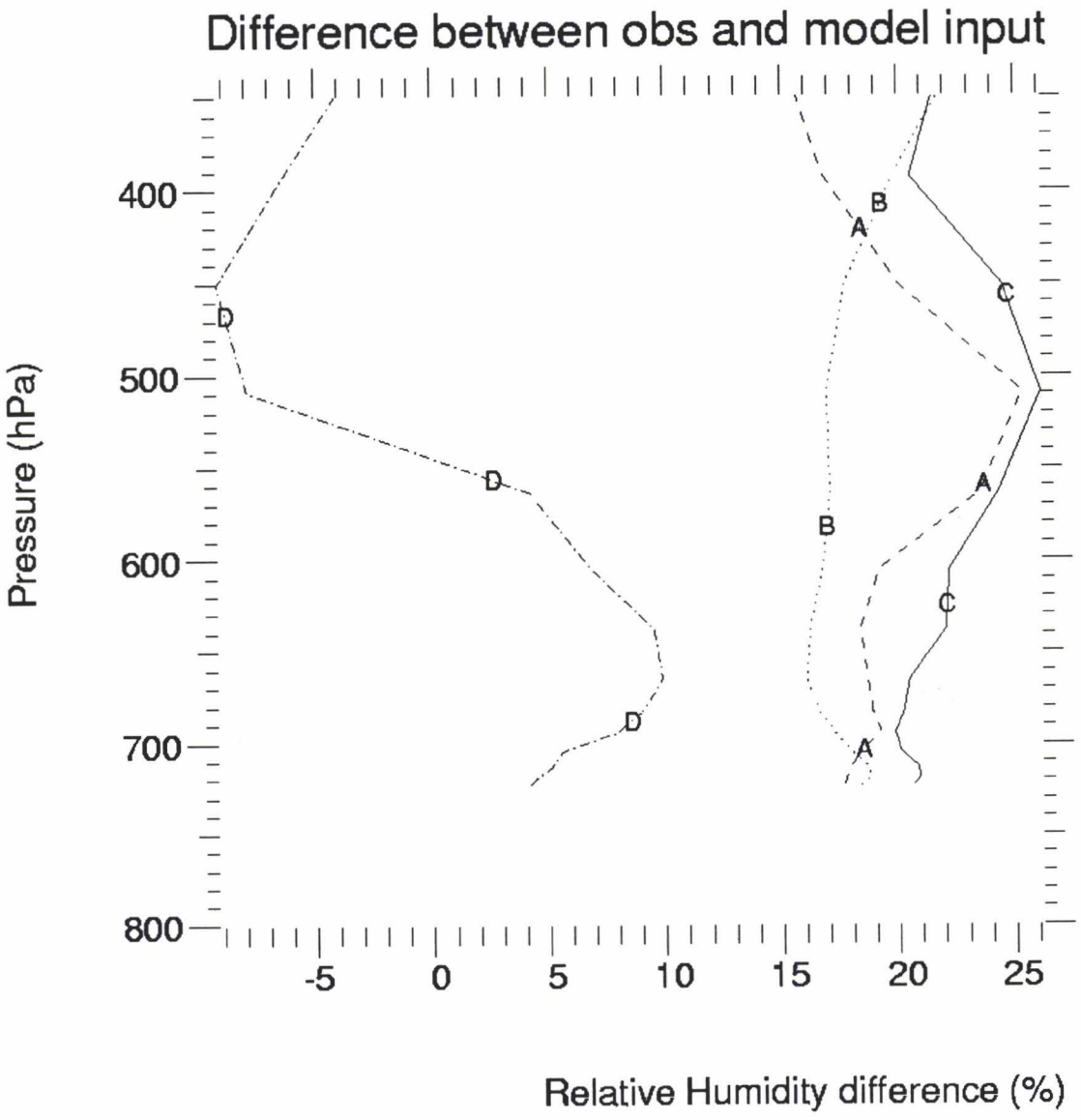

Figure 57. Average and RMS differences (SLC - input) from input dataset to SLC rawinsonde, as well as standard deviations for relative humidity. The grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and SLC. "D", average difference between input and SLC.

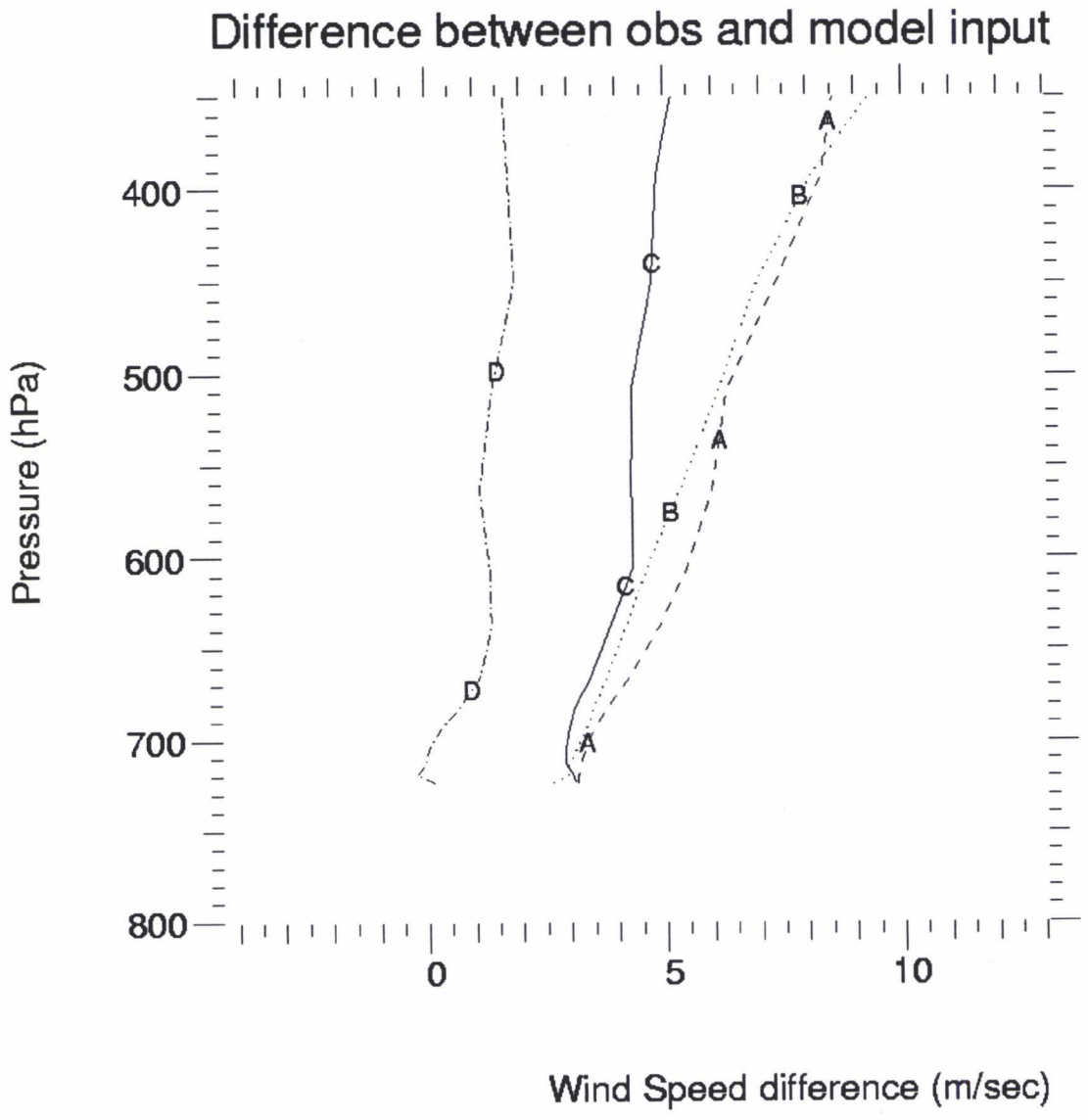

Figure 58. Average and RMS differences (SLC - input) from input dataset to SLC rawinsonde, as well as standard deviations for wind speed. The grid point closest to SLC was used to compare to the rawinsonde. "A", Salt Lake City rawinsonde dataset standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and SLC. "D", average difference between input and SLC.

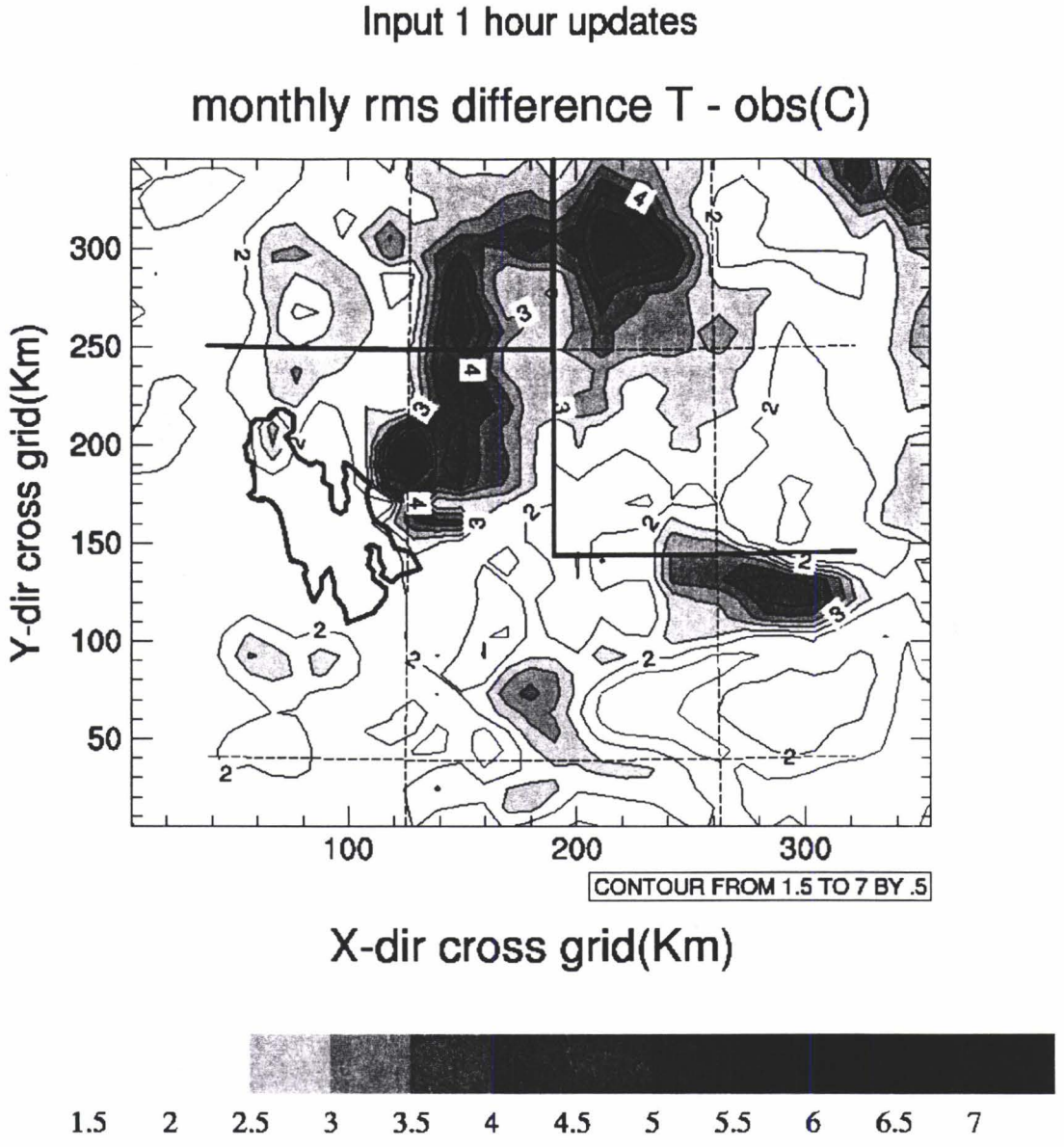

Figure 59. Contour plot of the RMS difference between input dataset and observed daily T_{ave} for September, 1981. Observed temperature calculated using the local polynomial regression technique on the mean of daily maximum and minimum temperatures. Input dataset temperature is calculated by finding the mean of the 6 hour data.

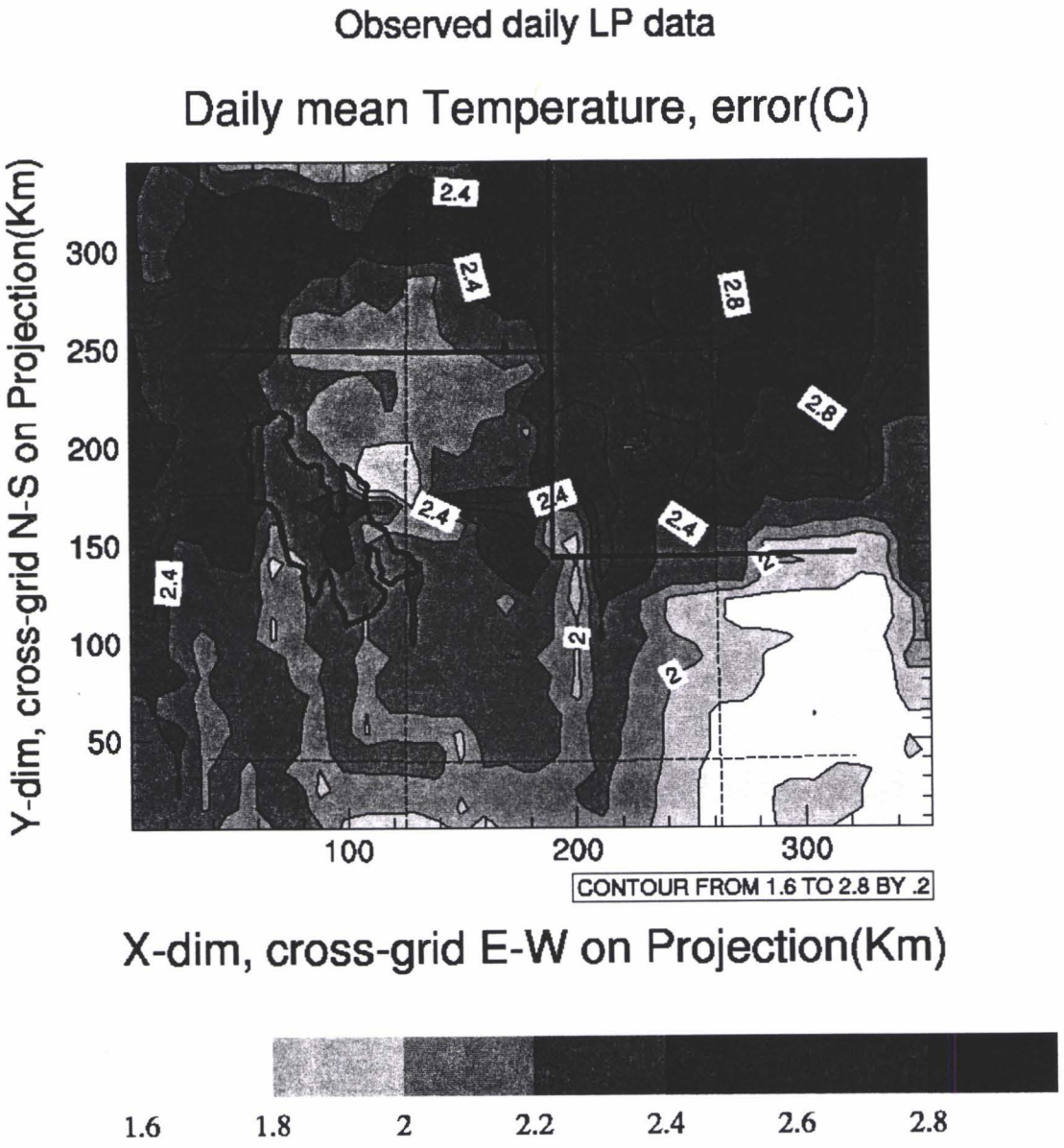

Figure 60. Average observed temperature error for September 1981. Observed error calculated using the local polynomial regression technique.

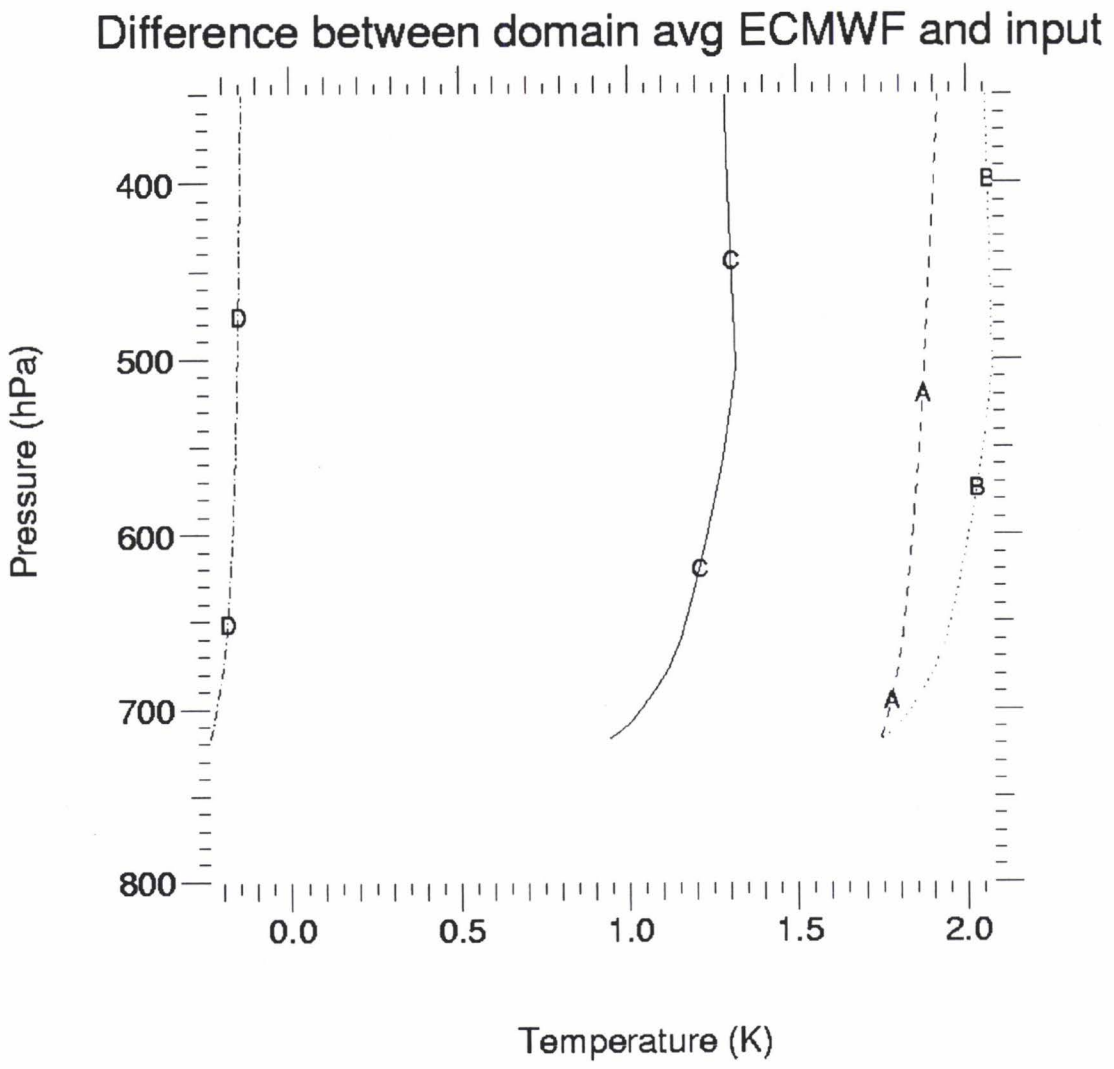

Figure 61. Average and RMS differences (ECMWF - input) of domain averaged temperature between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF.

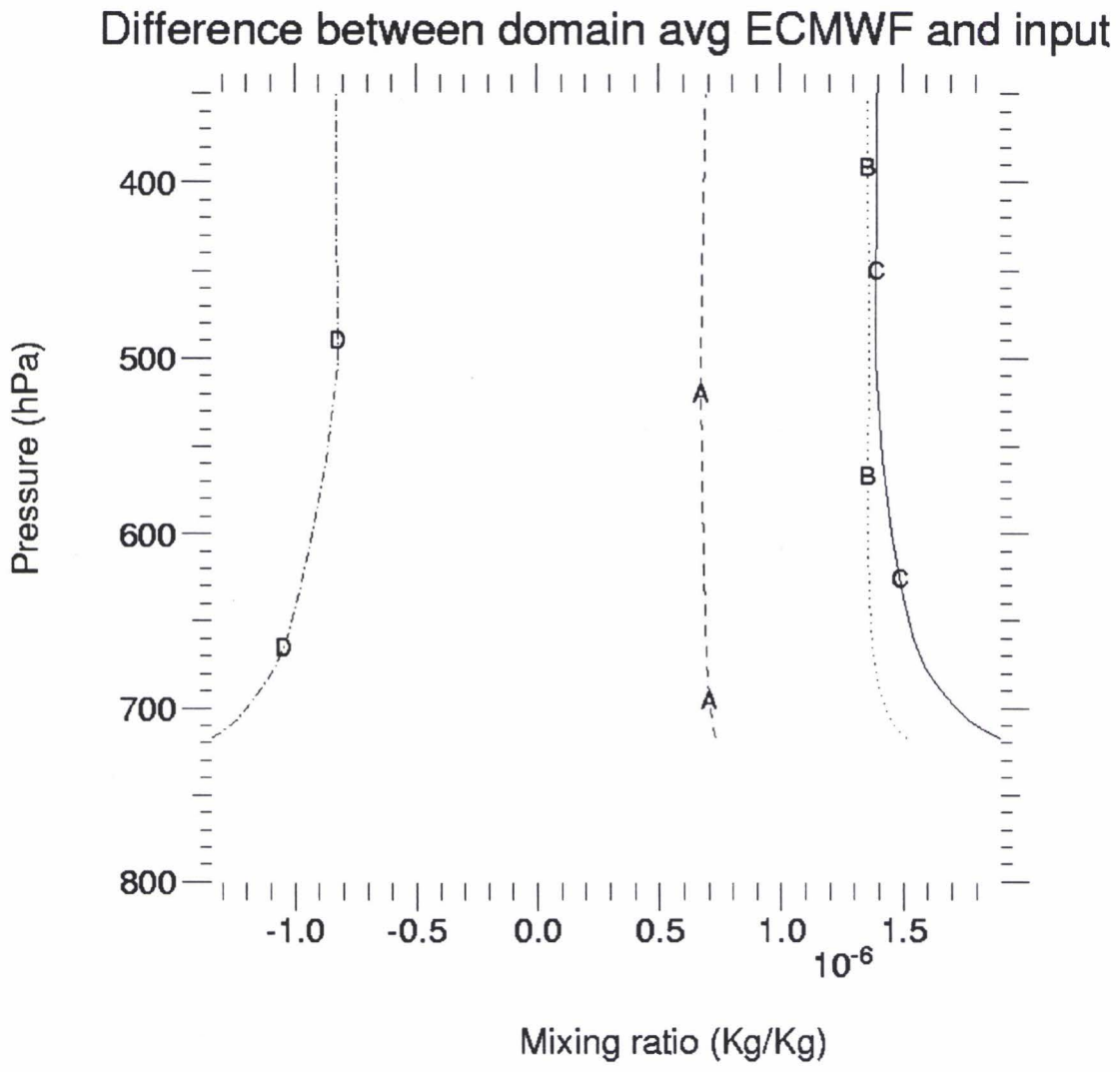

Figure 62. Average and RMS differences (ECMWF - input) of domain averaged water vapor mixing ratio between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF.

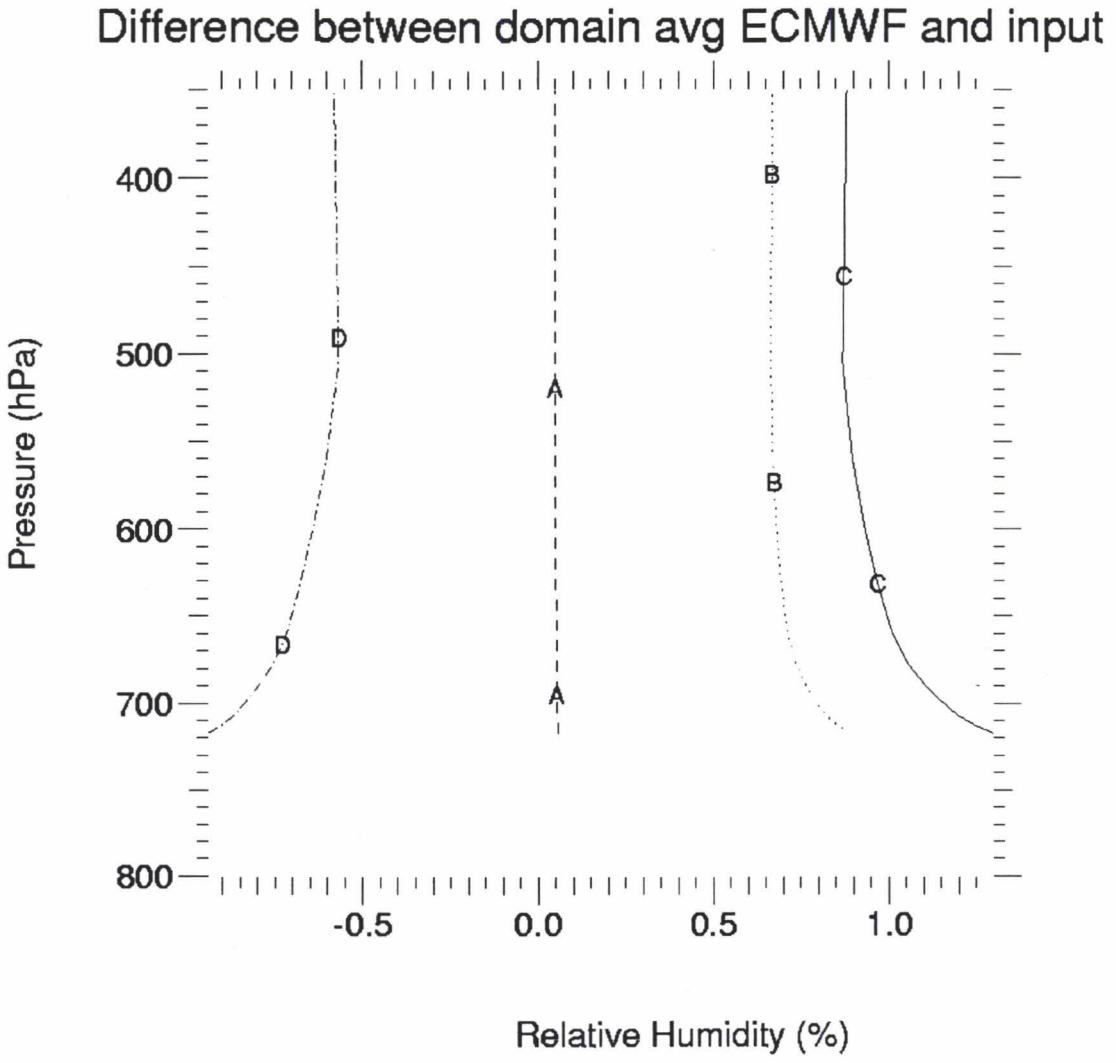

Figure 63. Average and RMS differences (ECMWF - input) of domain averaged relative humidity between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF.

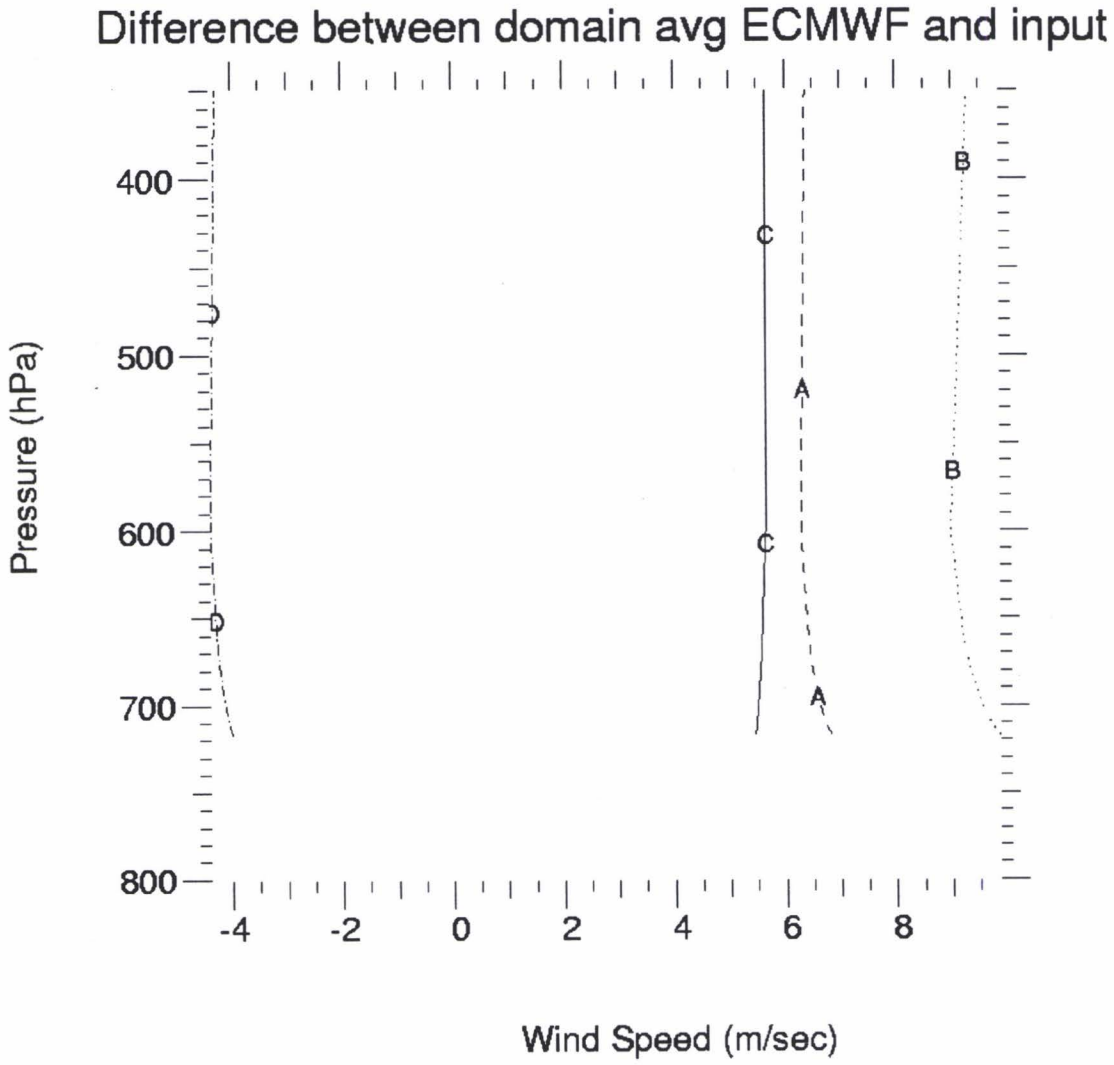

Figure 64. Average and RMS differences (ECMWF - input) of domain averaged wind speed between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF.

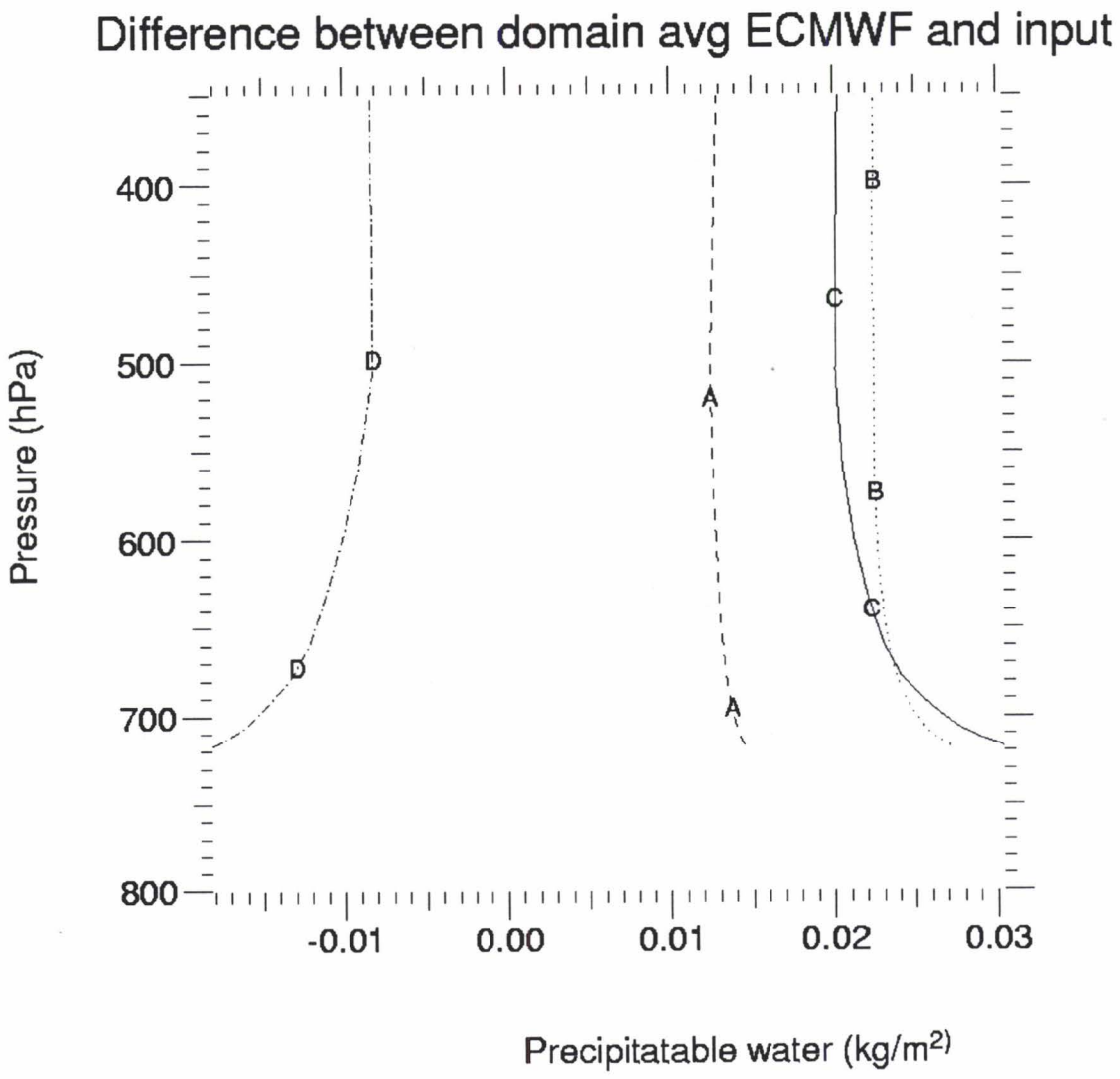

Figure 65. Average and RMS differences (ECMWF - input) of domain averaged total water mass between the ECMWF and input datasets, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. Only datapoints for which the elevation of the ECMWF dataset is within 3 m of the input dataset are considered in the domain average. "A", ECMWF standard deviation. "B" input dataset standard deviation. "C", RMS difference between input and ECMWF. "D", average difference between input and ECMWF.

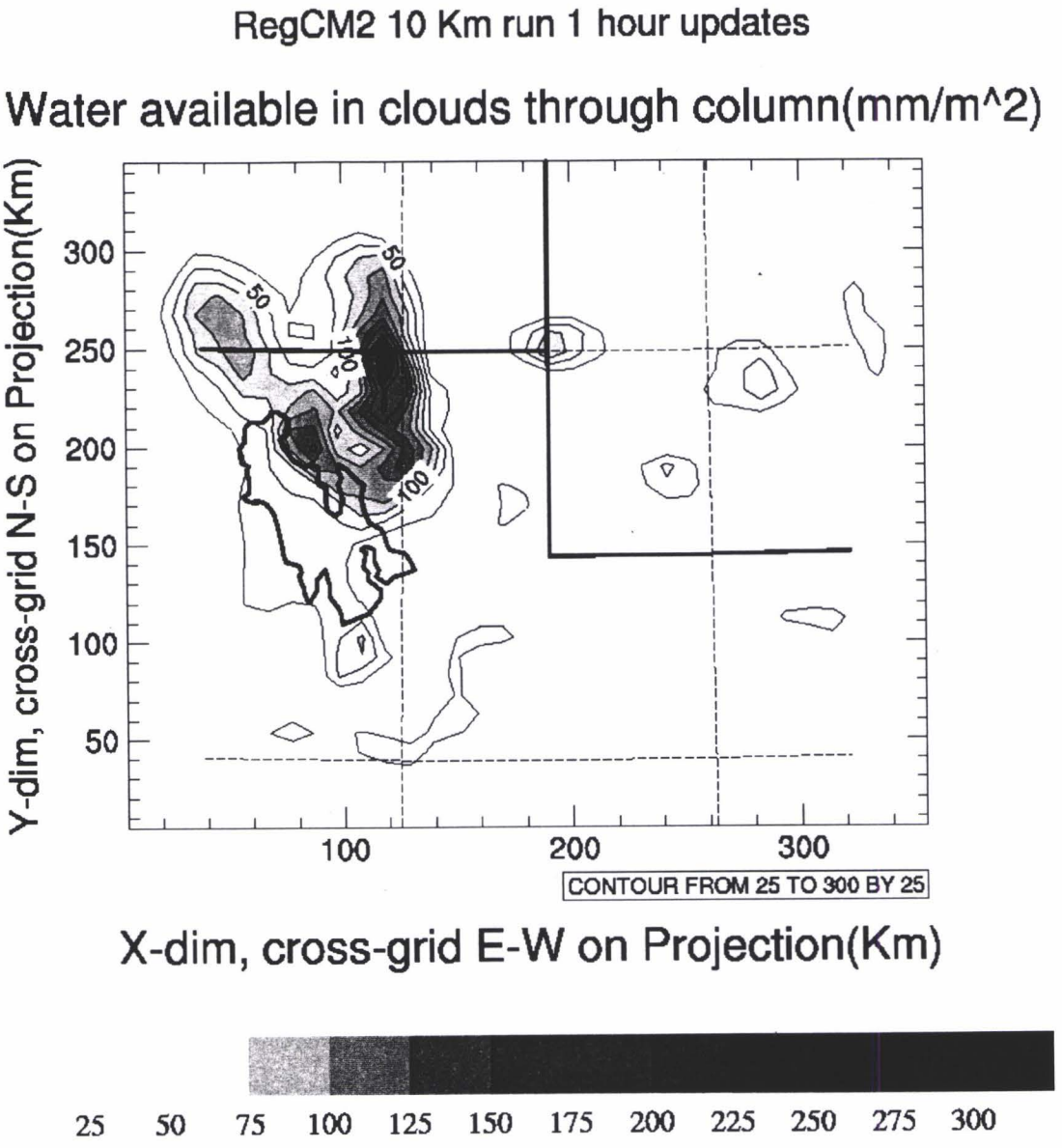

Figure 66. Control simulation total column cloud water totaled over September 1981 in mm of water per unit area.

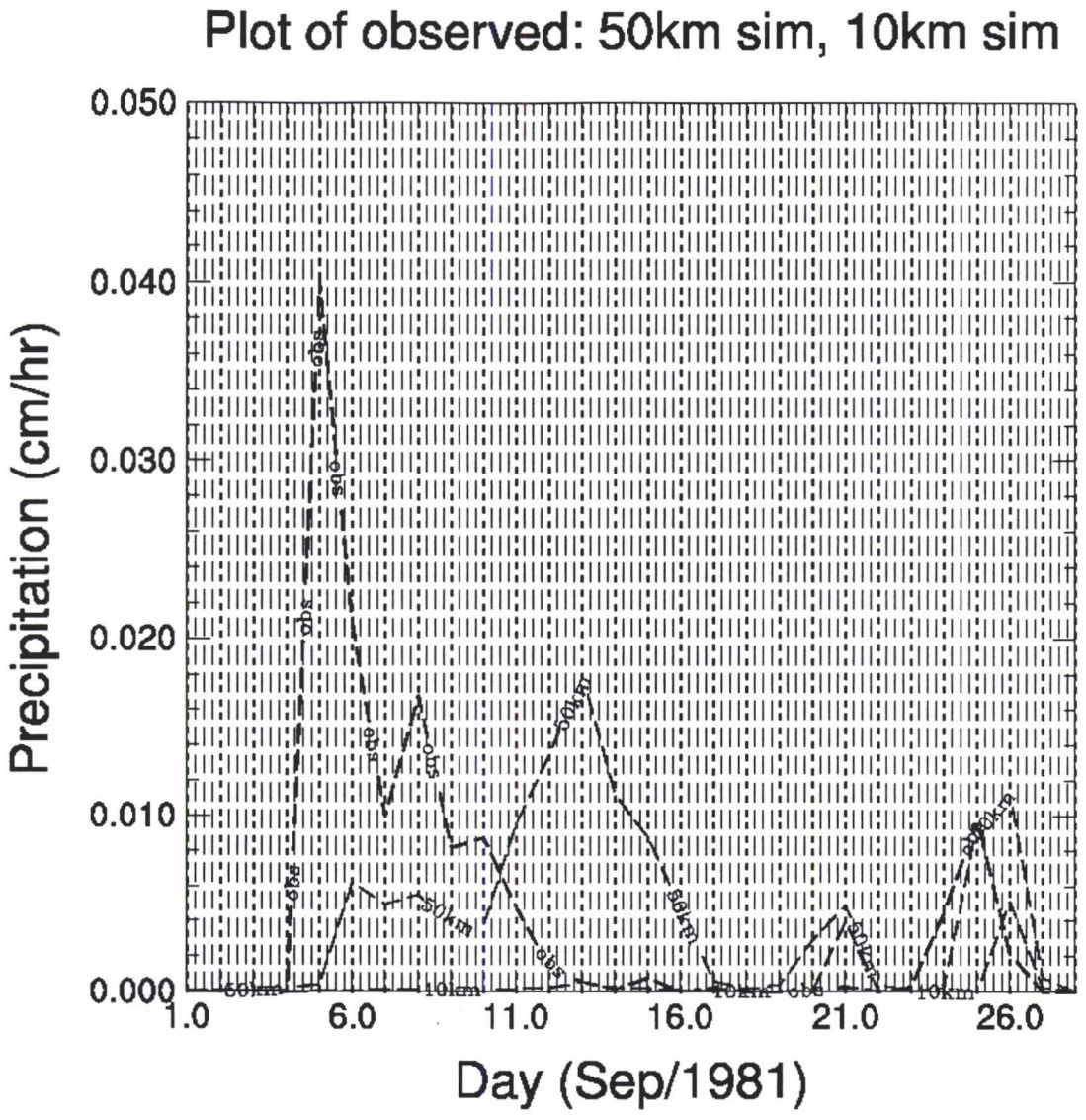

Figure 67. Comparison of domain averaged precipitation with time. Observed (black), average of 50 km simulation over the region (green), and average of the 10 km simulation (red). Dotted grid marks each day, and the dashed grid marks each 6 hour interval.

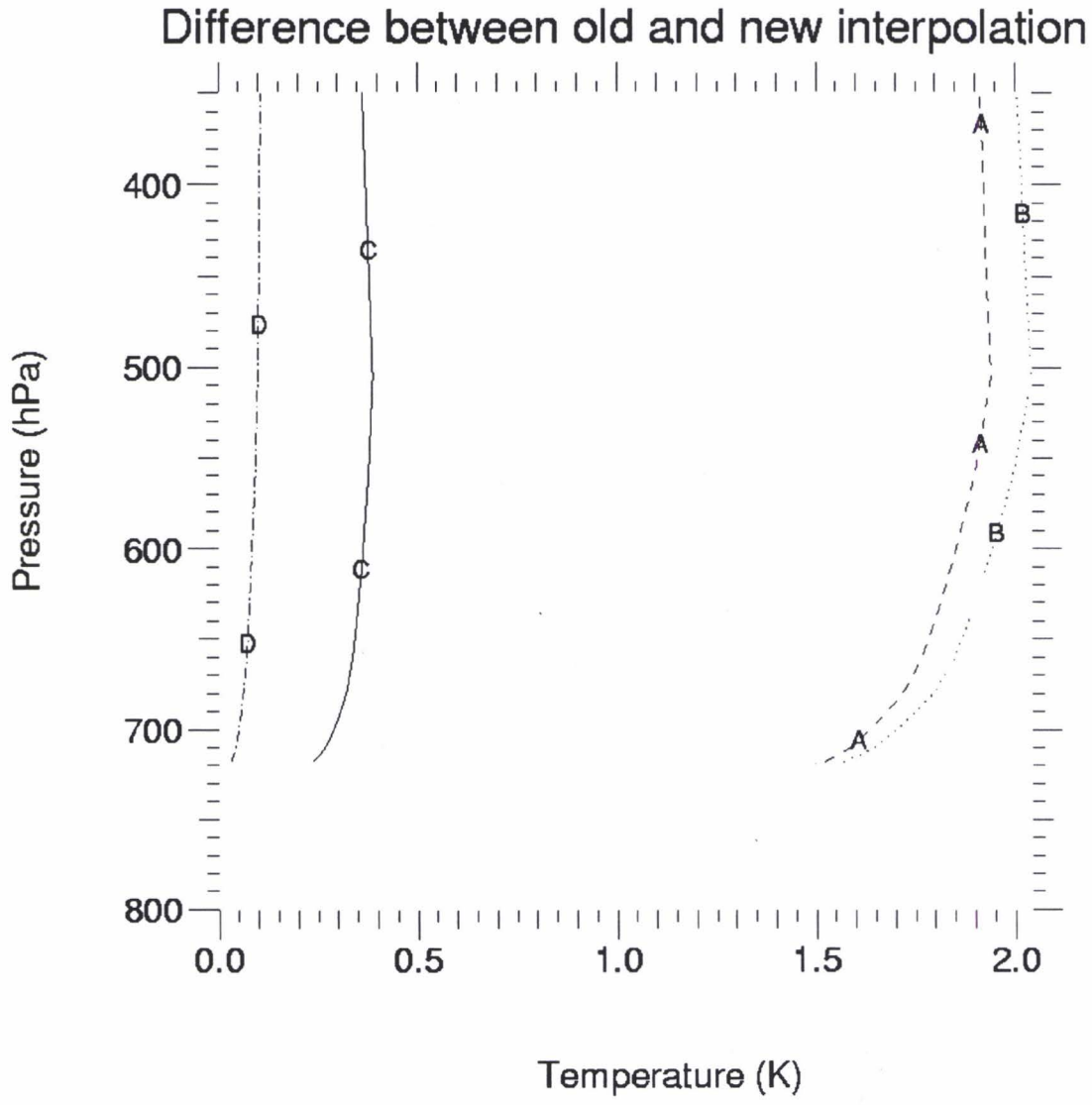

Figure 68. Average and RMS differences (new - old) between the domain average for the control input dataset and the new interpolation methods for temperature, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. "A", "new" standard deviation. "B" "old" dataset standard deviation. "C", RMS difference between "new" and "old". "D", average difference between "new" and "old".

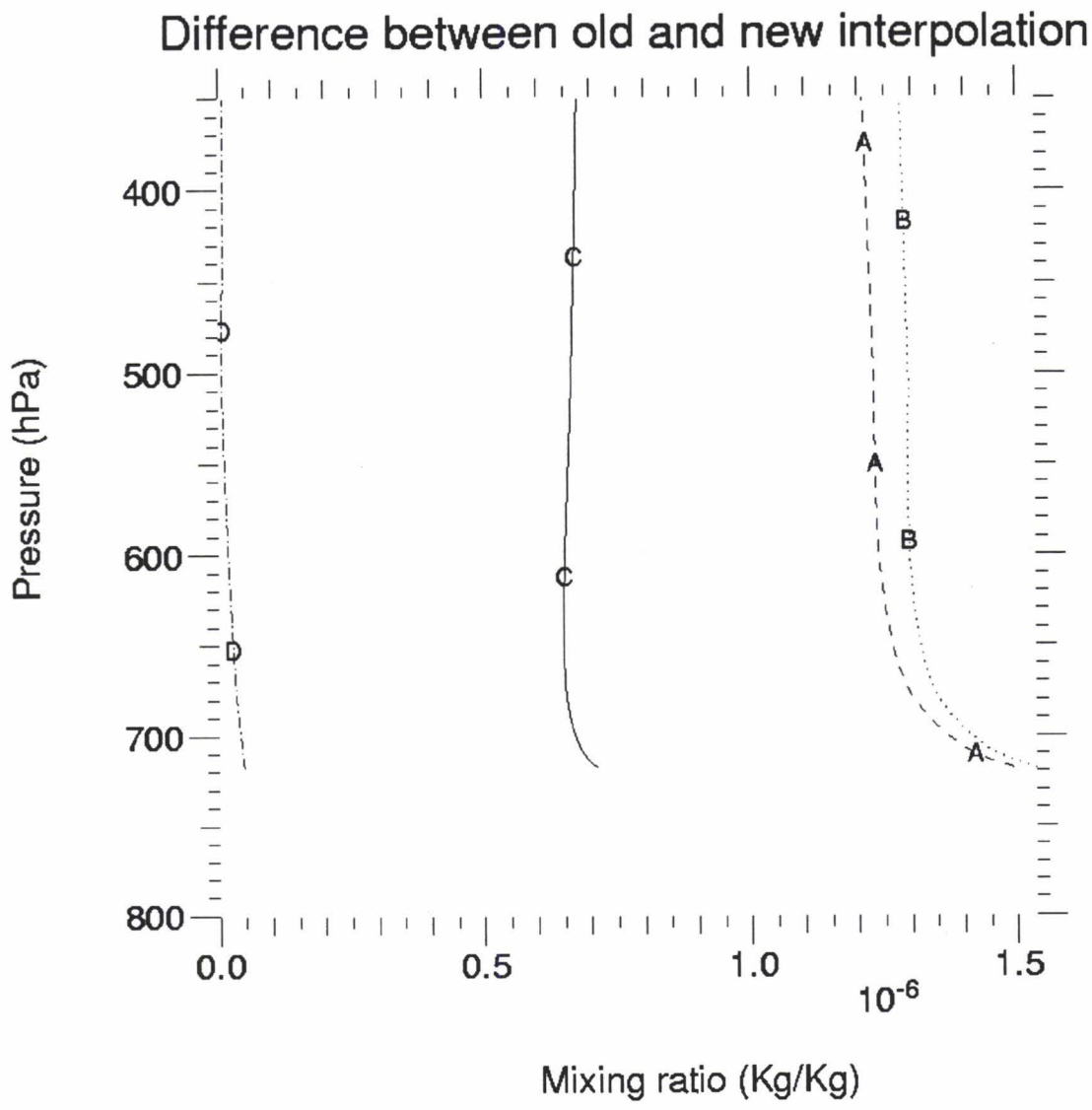

Figure 69. Average and RMS differences (new - old) between the domain average for the control input dataset and the new interpolation methods for water vapor mixing ratio, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. "A", "new" standard deviation. "B" "old" dataset standard deviation. "C", RMS difference between "new" and "old". "D", average difference between "new" and "old".

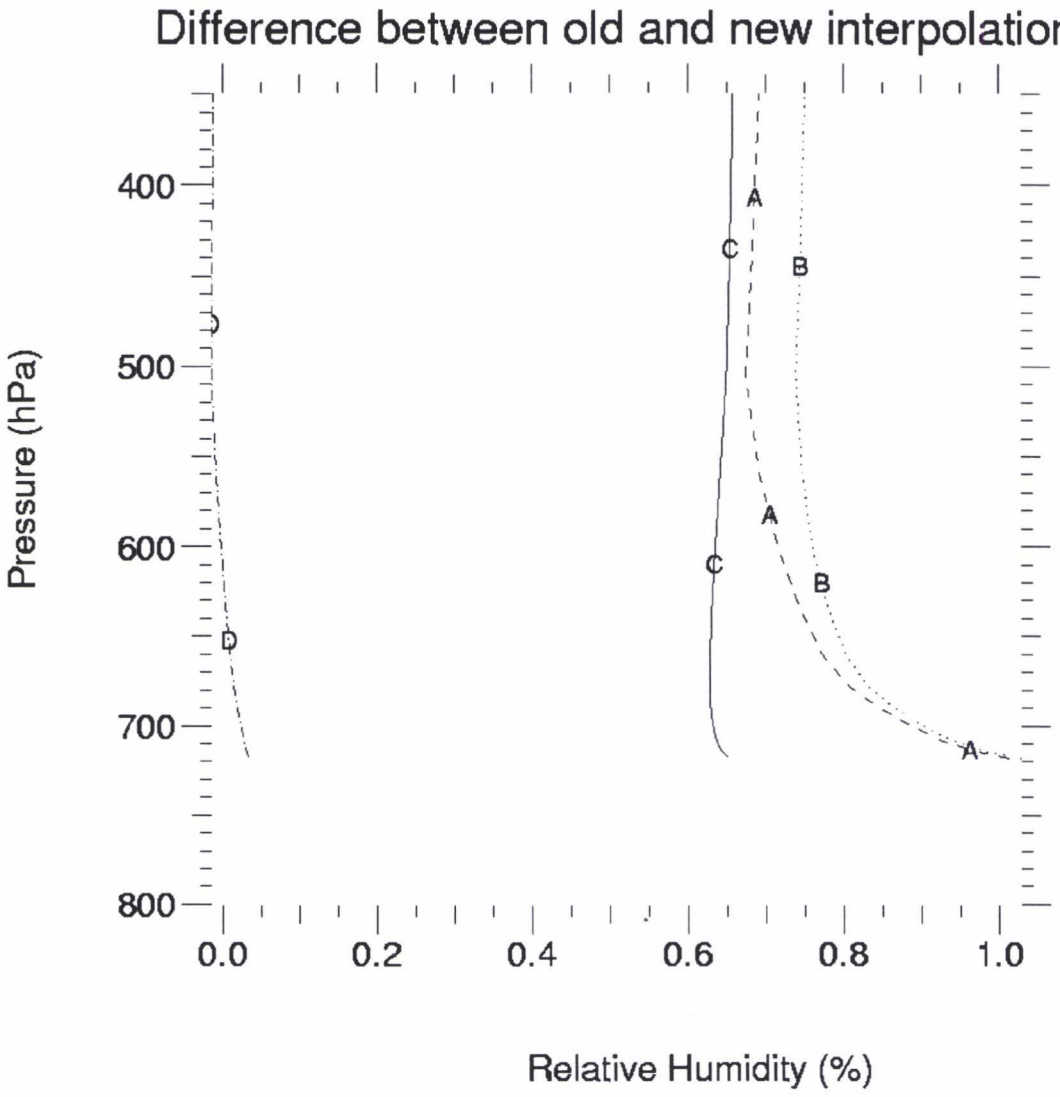

Figure 70. Average and RMS differences (new - old) between the domain average for the control input dataset and the new interpolation methods for relative humidity, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. "A", "new" standard deviation. "B" "old" dataset standard deviation. "C", RMS difference between "new" and "old". "D", average difference between "new" and "old".

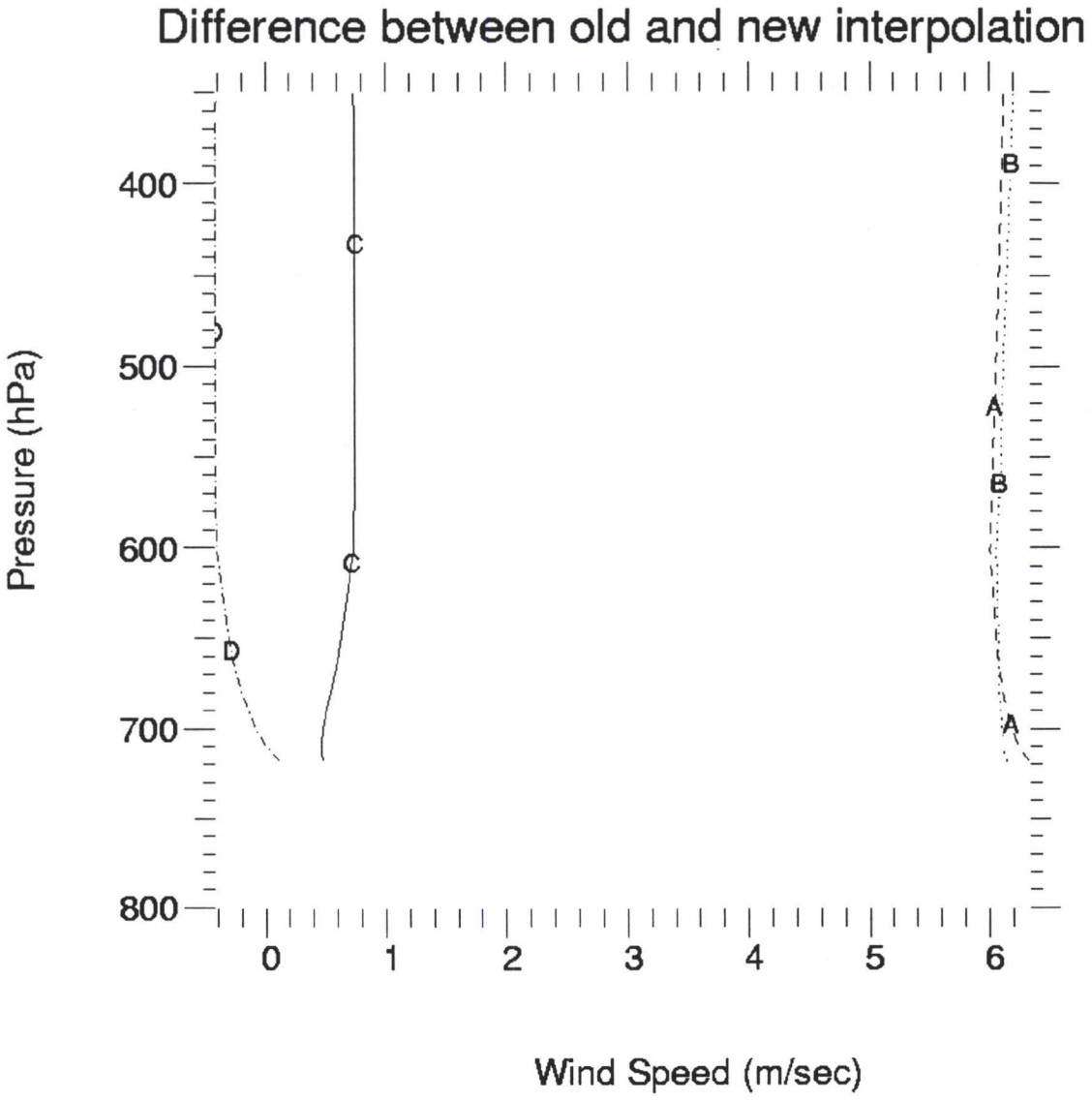

Figure 71. Average and RMS differences (new - old) between the domain average for the control input dataset and the new interpolation methods for wind speed, as well as standard deviations. Data is averaged over each sigma layer and the average pressure for that sigma layer given. "A", "new" standard deviation. "B" "old" dataset standard deviation. "C", RMS difference between "new" and "old". "D", average difference between "new" and "old".

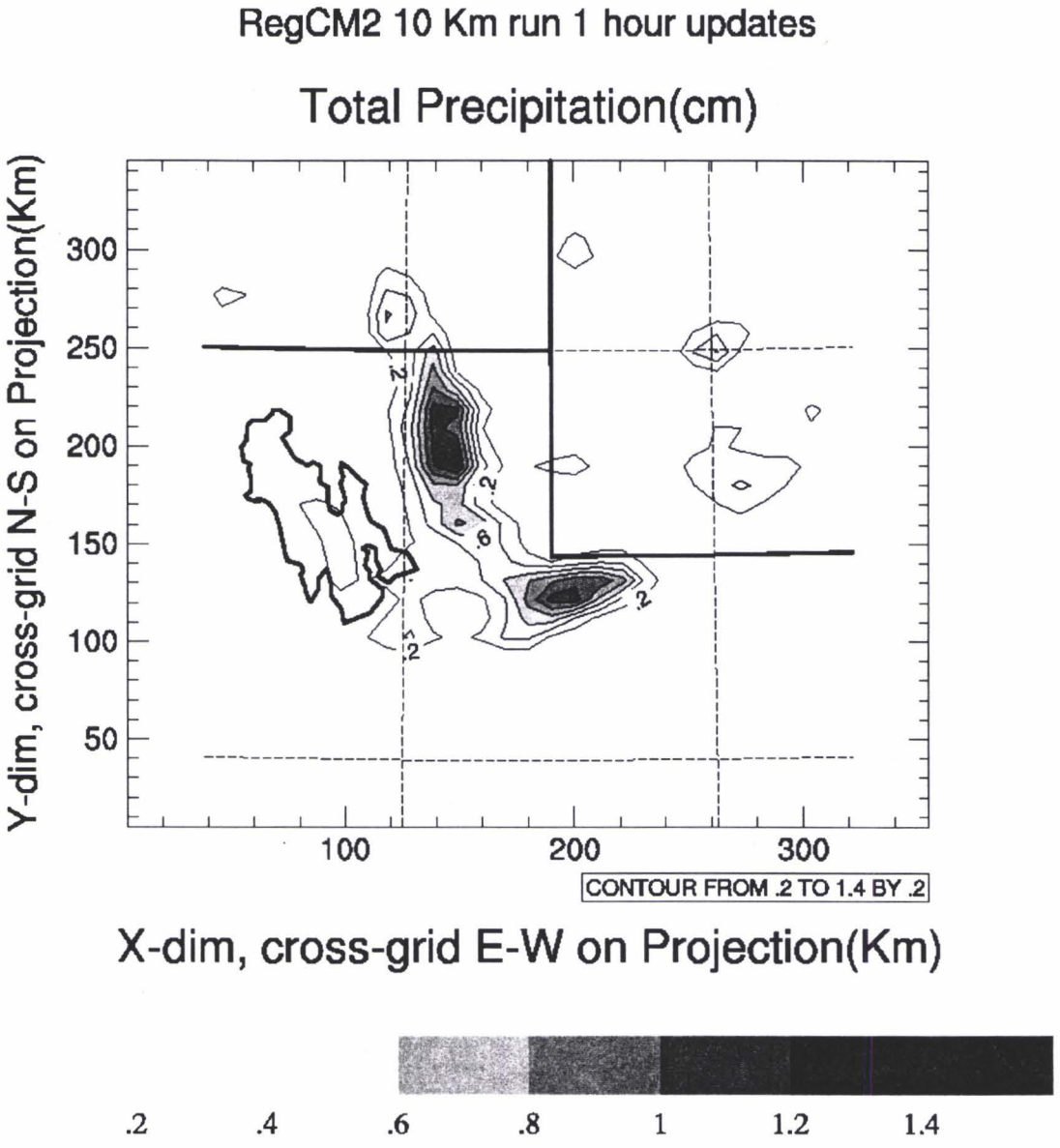

Figure 72. Precipitation for 10 km simulation with hourly LBC updates.

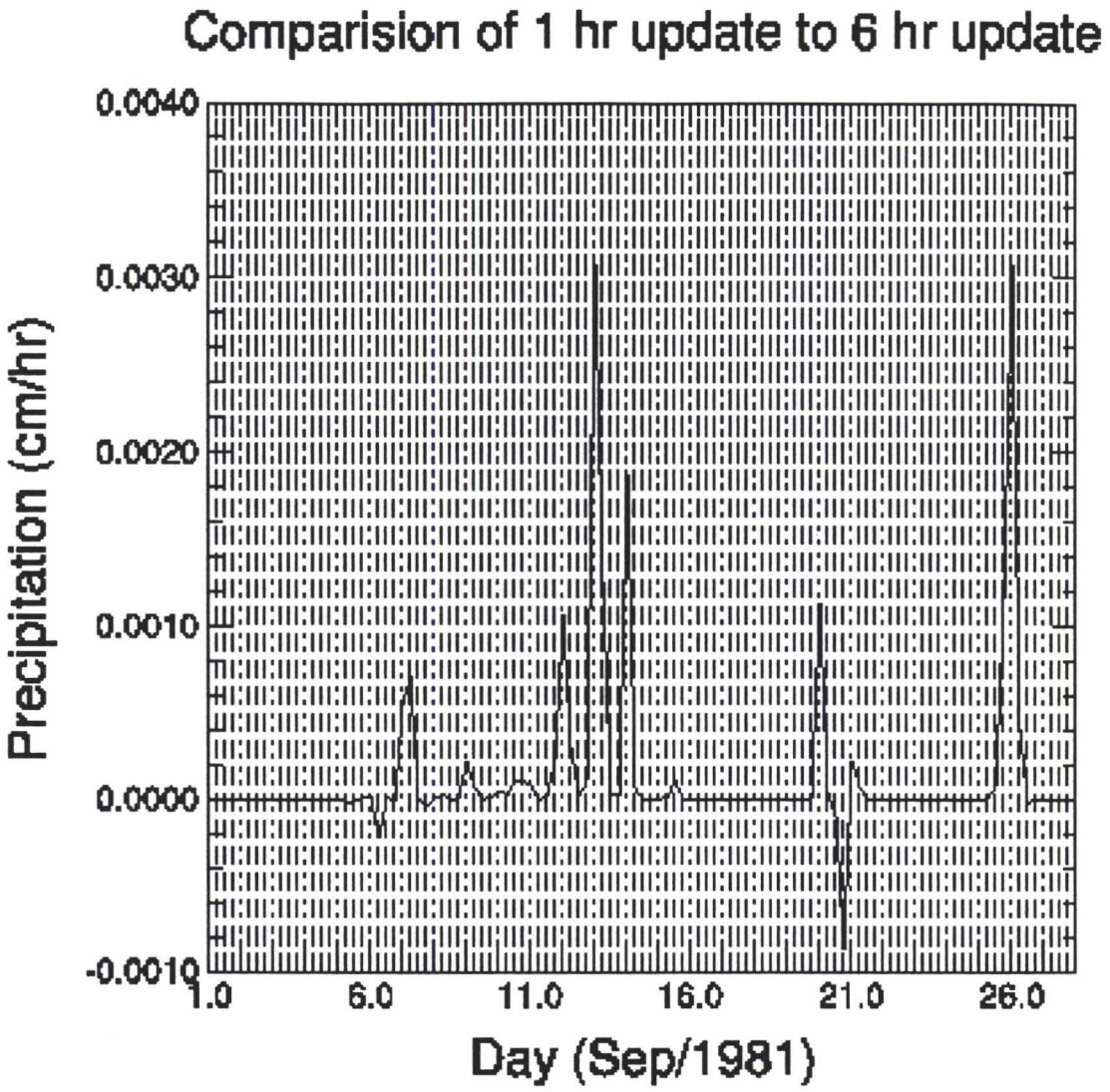

Figure 73. Comparison of grid-scale average precipitation from control to 1 hour update case. Dotted grid marks each day, and the dashed grid marks each 6 hour interval.

CHAPTER V: SUMMARY AND CONCLUSIONS

We review the performance of model simulations at both the 50 km and 10 km scales. We also discuss possible changes to improve performance at both scales, and how the changes in each scale might relate to each other. Some of the significant findings of our studies are as follows.

1. Model comparisons require carefully constructed gridded observed datasets that include error estimates.
2. RegCM2 50 km simulations show good model skill compared to observed anomalies.
3. RegCM2 diurnal surface temperature variation is too low.
4. Significant model bias exists in precipitation fields.
5. Convective precipitation events are poorly modeled by RegCM2.
6. Ten km simulations over northern Utah show precipitable water and precipitation at an order of magnitude too low.
7. Using explicit cloud and rain water seems to be important for 10 km simulations as well as including cloud and rain water on the LBC.
8. Explicit cloud water terms need to be adjusted for the 10 km domain. This is because 10 km vertical velocities will be smaller than those used by convective parameterization at coarser resolutions.
9. Differences in the preparation of LBC datasets and their update frequency can affect model results.

In chapter III we discuss the performance of RegCM2 for our 50 km simulations over the western US. The model's precipitation anomalies compare very well to the observed. The model skill also shows that the bias-corrected model performs better than a climatological forecast in almost all areas. The wintertime threat score for the zero- level precipitation threshold also shows that the model accurately predicts if precipitation will occur. However, the model predicts substantially more precipitation than observed in almost all areas.

From the summertime simulations it is obvious that the model performs very poorly in respect to predicting convective precipitation. The simulation is also very poor in the southern regions where convective precipitation plays a dominant role. Thus the convective parameterization of the model needs tangible improvement to refine performance. From the intensity distributions it appeared that the model produces storms of about the right intensity. The wet-spell/dry-spell frequency distributions showed that the over prediction of precipitation seemed to be due to the fact that storms are triggered too often. This likely will require adjustment to the model's precipitation physics – both the grid-scale resolved precipitation and the convective precipitation parameterization.

In chapter IV we discuss the 10 km simulations. One of the recommendations is to do nested 10 km simulations with one of the required input fields being cloud water and rain water. We also

suggest that running the 50 km simulation with explicit cloud and rain water is likely to improve the performance of the 50 km simulations. Although this is unlikely to change the poor performance of the convective precipitation, explicit cloud and rain water may improve grid-scale resolvable precipitation. An advantage of using the explicit cloud and rain water parameterization is that it has tuning parameters that can be adjusted to improve the resultant simulation. As discussed in chapter IV, the auto-conversion threshold and rate can be slightly adjusted to lower the grid-scale precipitation.

The poor performance of the convective parameterization is a known problem, with many different attempts at providing better solutions [*Molinari and Dudek*, 1992]. It is unlikely that simple adjustments to the parameterization might yield better results. The main problem is that storms are triggered too often. This can be because the potential temperature profiles indicate instabilities too frequently or the large-scale triggering mechanisms in the convective parameterization are not functioning correctly. Perhaps too much water is being transported from the LBC and/or the Pacific Ocean, or there are problems with the vertical profile of temperature due to weaknesses in the boundary layer or surface layer physics. Analysis of the problems with convective precipitation, much less solutions, will be quite difficult. Our 50 km simulation also points out the need for longer term simulations. The bias for any single month can differ significantly from the long-term bias. However, 1- to 3-year simulations might be sufficient for tuning exercises.

The 10 km simulation displays a different behavior with the grid-scale average precipitation being too small and having unrealistically high concentrations of precipitation on up-slopes. Moreover, convective precipitation seems to be a nonissue with only a very small percentage of the total precipitation coming from convective precipitation. One of the main problems with the simulation is that the total amount of water in all forms moving into the system is too low. Since the implicit form of cloud and rain water in the 50 km simulation makes it difficult to advect it into the system, we ignore this component (as others have done) and let the system develop its own storms. However, the 10 km grid domain is too small for sufficient time to develop substantial storm systems. The 50 km grid domain, being much larger and having a substantial amount of ocean in the system, has time to develop storm systems. The other problem seems to be that too much water is locked up in the clouds. The cloud physics need to be tuned so that clouds do not have to be so thick to form precipitation.

An alternative solution to the one-way nesting scheme that we demonstrate here is to use a different model that has two-way nesting to higher resolution grids in the same model simulation (such as MM5 or RAMS). These models eliminate the problems (LBC update frequencies that are too low) we outline in this study by updating the LBC for the higher resolution domain every time-step. The cost for this of course is substantial and increases the amount of computing time required to perform such a simulation. It also lacks the ability to nest to a different area for another high-resolution simulation without running the course resolution over again. Because of this, one-way nesting still retains its utility. As GCM resolutions increase, the need to have the ability to one-way nest high-resolution GCM simulations to even higher resolution regional climate models will remain.

The need for using the type of technology developed in chapter II for model comparison will remain high. As we find in chapters III and IV, error estimates are critical in making assessments

on the model. In chapter II we find that monthly totals of the observed data computed daily have large errors and a noticeable bias from observed data computed monthly. Therefore, the monthly totals we use in chapter III and IV are computed from monthly data. In chapter IV we also need the error estimates to judge the possibility and the likely offset that up-slopes receive more precipitation. Thus our work is an important contribution to the modeling community.

We are left to conclude that meteorological variable input driving today's mesoscale models do not provide a significantly improved database for high spatial resolution, terrain mapped Forest Service application modeling of either erosion or plant and animal growth. The next generation models now being used for heat island and surface hydrology studies allow for internal grid data inputs so that observed rainfall and temperature values can be injected in to the simulation. These inputs help control storm phasing and drive short term surface layer temperatures and precipitation toward more reasonable levels. However, the "too wet" and "too often" convective precipitation phenomenon that still exists in RegCM2 after all our effort strongly suggests that reasonable answers cannot be found in "carefully parameterized" hydrostatic simulations. While computer power and data storage capability has increased by an order of magnitude in the last 5 years, the requirements of conducting long term simulations with non-hydrostatic codes still exceed the computer capabilities of non-global climate simulation sized research centers. Given another generation or two of computer capability and model development, it may well become practical to link the meteorological and climatological data sets into boundary conditions that provide sufficient constraint to the model to allow valuable temperature and precipitation simulations that do meet Forest Service requirements in complex terrain. However, the current state of the art lies with the surface observation extrapolation schemes, such as the local polynomial scheme utilized in Chapter II.

References

Molinari, J., and M. Dudek, Parameterization of convective precipitation in mesoscale numerical models: A critical review, *Mon. Weather Rev.*, *120*, 326-344, 1992.